

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису
УДК 004.942:519.216.3

До захисту допущено
В. о. завідувача кафедри ММСА

О.Л.Тимошук

«___» _____ 2020 р.

Магістерська дисертація

на здобуття ступеня магістра за спеціальністю 122 Комп'ютерні науки
на тему: «Система для прогнозування нелінійних нестационарних процесів в
екології та економіці»

Виконав:

студент II курсу, групи КА-93 мп
Тертичний Роман Віталійович

Керівник: професор кафедри ММСА,
д.т.н., професор,
Бідюк П. І.

Рецензент: декан ФІОТ
КПІ ім. Ігоря Сікорського
д.т.н., професор,
Теленик С.Ф.

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів
без відповідних посилань
Студент _____

Київ
2020

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)
Спеціальність — 122«Комп'ютерні науки»

ЗАТВЕРДЖУЮ

В. о. завідувача кафедри ММСА

О. Л. Тимошук

«___» _____ 2020 р.

ЗАВДАННЯ

на магістерську дисертацію студенту Тертичному Роману Віталійовичу

1. Тема дисертації: «Система для прогнозування нелінійних нестационарних процесів в екології та економіці», науковий керівник дисертації Бідюк Петро Іванович, д.т.н., професор, затверджені наказом по університету від «02» листопада 2020р. № 3182-с.

2. Термін подання студентом дисертації: 14 грудня 2020 р.

3. Об'єкт дослідження: нелінійні нестационарні процеси в екології та економіці, представлені статистичними даними стосовно їх розвитку

4. Предмет дослідження: ймовірно-статистичні та методи інтелектуального аналізу даних для моделювання і прогнозування розвитку нелінійних нестационарних процесів; множини критеріїв для аналізу адекватності моделей та оцінювання якості прогнозів

5. Перелік завдань, які потрібно розробити:

1) дослідити сучасний стан та особливості економічних та екологічних процесів у світі;

2) проаналізувати існуючі методи прогнозування економічних та екологічних процесів та виконати огляд використовуваних комп'ютерних систем для прогнозування;

3) дослідити методи та підходи до моделювання та прогнозування нелінійних та нестационарних процесів;

4) пошук даних для застосування в програмі;

5) вивчити та зрозуміти методики моделювання та прогнозування нелінійних нестаціонарних процесів та їхню структуру;

6) виконати побудову математичних моделей на реальних статистичних даних та отримати практичні результати, виконати аналіз результатів;

7) розробити стартап-проект виведення на ринок результатів дослідження;

8) розробити концептуальні висновки за результатами наукового дослідження

6. Орієнтовний перелік графічного (ілюстративного) матеріалу:

1). Функціональна блок-схема системи (рис.);

2). Інформаційні графіки (рис.);

3). Приклади результатів побудованих моделей (рис.);

4). Таблиці у розділі стартап-проекту та порівняльні таблиці результатів.

7. Дата видачі завдання: 01 вересня 2020 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації
1.	Концептуальний вступ дисертації. Формулювання об'єкта, предмета, цілі, завдань, новизни, практичної значущості результатів	06.09.2020 — 13.09.2020
2.	Перший розділ. Огляд літературно-інформаційних джерел. Характеристика існуючих проблем	14.09.2020 — 27.09.2020
3.	Другий розділ. Огляд математичних методів прогнозування нелінійних нестаціонарних процесів	28.09.2020 — 11.10.2020
5.	Третій розділ. Формування структури моделі прогнозування та основних її етапів	12.10.2020 — 21.10.2020
6.	Четвертий розділ. Реалізація системи та отримання практичних результатів	22.10.2020 — 06.11.2020
7.	П'ятий розділ. Стартап-проект	07.11.2020 — 14.11.2020
8.	Концептуальні висновки. Перспективи розвитку отриманих рішень	15.11.2020 — 22.11.2020

Студент

Р.В. Тертичний

Науковий керівник дисертації

П.І. Бідюк

РЕФЕРАТ

Магістерська дисертація: 129 с., 27 рис., 34 табл., 1 додаток, 16 джерел.

ЧАСОВІ РЯДИ, РЕГРЕСІЙНІ МОДЕЛІ, НЕЙРОННІ МЕРЕЖІ,
ПРОГНОЗУВАННЯ, ЕКОНОМІЧНІ ТА ЕКОЛОГІЧНІ ПРОЦЕСИ

Об'єктом дослідження є нелінійні нестационарні процеси в економіці та екології, представлені у вигляді часових рядів на основі статистичних даних стосовно їхньої динаміки.

Предметом дослідження є ймовірнісно-статистичні методи та методи інтелектуального аналізу даних для моделювання і прогнозування розвитку нелінійних нестационарних процесів, а також множини критеріїв для аналізу адекватності моделей та оцінювання якості прогнозів.

Метою дослідження є аналіз характеру поведінки динамічних процесів економіки та екології на основі часових рядів та прогнозування за допомогою різних видів авторегресійних моделей та нейронних мереж, а також реалізація системи моделювання у вигляді програмного продукту з отриманими емпіричними результатами для порівняння адекватності моделей, оцінки прогнозу та візуалізації досліджуваних процесів.

Методами дослідження моделювання та прогнозування нелінійних нестационарних процесів є методи регресійного аналізу та методи інтелектуального аналізу даних.

Новизною є підвищення якості моделювання і прогнозування нелінійних нестационарних процесів в екології та економіці.

ABSTRACT

Master's thesis: 129 pages, 27 images, 34 tables, 1 append., 16 sources.

The theme: «System for forecasting nonlinear and nonstationary processes in economics and ecology ».

TIME SERIES, REGRESSION MODELS, NEURAL NETWORK, FORECASTING, ECONOMIC AND ECOLOGICAL PROCESSES

The object of the study is nonlinear non-stationary processes in economics and ecology, presented in the form of time series based on statistics on their dynamics.

The subject of the research is probabilistic-statistical methods and methods of data mining for modeling and forecasting the development of nonlinear non-stationary processes, as well as many criteria for analyzing the adequacy of models and assessing the quality of forecasts.

The purpose of the study is to analyze the behavior of dynamic processes of economics and ecology based on time series and forecasting using different types of autoregressive models and neural networks, as well as the implementation of a modeling system as a software product with empirical results to compare models, forecast and visualize processes.

Methods of research of modeling and forecasting of nonlinear nonstationary processes are methods of regression analysis and methods of data mining.

The novelty is to improve the quality of modeling and forecasting of nonlinear nonstationary processes in ecology and economics.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ І ПОЗНАЧЕНЬ	9
ВСТУП	10
РОЗДІЛ 1	12
ОСОБЛИВОСТІ МОДЕЛЮВАННЯ ПРОЦЕСІВ В ЕКОНОМІЦІ ТА ЕКОЛОГІЇ, ІСНУЮЧІ РЕЗУЛЬТАТИ	12
1.1 Аналіз проблематики сучасних економічних та екологічних процесів	12
1.1.1 Проблеми екологічних процесів у світі.....	13
1.1.2 Проблеми економічних процесів у світі.....	19
1.2 Існуючі методи прогнозування та особливості економічних та екологічних процесів	23
1.3 Огляд деяких комп'ютерних систем для побудови моделей прогнозування та аналізу даних	27
Висновки до розділу 1 і постановка задачі.....	32
РОЗДІЛ 2	34
МЕТОДИ ТА ПІДХОДИ ДО МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ	34
2.1 Побудова авторгресійних моделей AR, ARMA, ARMAX.....	34
2.2 Прогнозування на основі нейронних мереж	40
2.2.1 Нейронні мережі. Рекурентна нейронна мережа.....	41
2.2.2 Нейромережа з довгою короткостроковою пам'яттю (LSTM)	44
Висновки до розділу 2	47
РОЗДІЛ 3	49
МЕТОДИКА МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ І ЇЇ СТРУКТУРА	49
3.1 Формування структури моделі	49
3.1.1 Аналіз нелінійності.....	51
3.1.2 Перевірка на стаціонарність	52

3.1.3 Наявність коінтегрованості.....	55
3.1.4 Аналіз на гетероскедастичність	57
3.2 Критерії вибору кращої моделі та прогнозу.....	60
3.2.1 Критерії адекватності моделі.....	61
3.2.2 Оцінювання точності прогнозу	63
Висновки до розділу 3	65
РОЗДІЛ 4	66
ПОБУДОВА МАТЕМАТИЧНИХ МОДЕЛЕЙ ЕКОНОМІЧНИХ ТА ЕКОЛОГІЧНИХ ПРОЦЕСІВ	66
4.1 Програмна реалізація та її архітектура	66
4.2 Аналіз вибору інструментальної платформи для реалізації системи	67
4.3 Побудова математичних моделей та прогнозування на основі статистичних даних	69
4.3.1 Побудова моделей на основі даних ВВП	69
4.3.2 Побудова моделей на основі даних забруднення повітря	79
Висновки до розділу 4	85
РОЗДІЛ 5	86
РОЗРОБКА ВЛАСНОГО СТАРТАП ПРОЕКТУ	86
5.1 Сутність та особливості стартап проектів	86
5.2 Формування команди стартапу	89
5.3 Розробка продукту для стартапу	91
5.4 Розроблення бізнес-моделі стартапу	95
5.5 Маркетингове планування стартапу	97
5.6 Бізнес-план стартап проекту	102
5.7 Патентний пошук	105
5.8 Правові аспекти реалізації стартапів, інтелектуальна власність та патентування	106

5.9 Інвестиційна та фінансове забезпечення стартап-проекту	120
Висновки до розділу 5	123
ВИСНОВКИ ПО РОБОТІ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ.....	124
ПЕРЕЛІК ПОСИЛАНЬ.....	126
ДОДАТОК А ЛІСТИНГ ПРОГРАМИ	129

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ І ПОЗНАЧЕНЬ

AR(AP) – авторегресійне рівняння

ARMA(APKC) – авторегресійне рівняння з ковзним середнім

ARIMA(APIKC) – авторегресійне рівняння з інтегрованим ковзним середнім

SES(E3) – експоненціальне згладжування

DW – критерій Дарбіна-Уотсона

ACF(AКФ) – автокореляційна функція

PACF(ЧАКФ) – часткова автокореляційна функція

ME(KC) – ковзне середнє

SSE(СКП) – сума квадратів похибок моделі

СП – середня похибка прогнозу

СПП – середня похибка в процентах

RMSE – середньоквадратична похибка

MAPE(СаПП) – середня абсолютна похибка у процентах

R^2 – коефіцієнт детермінації

AIC – інформаційний критерій Акайке

Theile – коефіцієнт Тейла

LSTM – Long short-term memory

ВСТУП

В умовах сучасності для успішної конкурентоспроможної діяльності на ринку і розвитку економіки та екології країни в цілому необхідно використовувати нові принципи управління, адекватні рівню сучасної економіки та екології. Прогнозування як напрямок економічних та екологічних наук, орієнтоване на оптимізацію ідеології управління, в повній мірі відповідає цілям і задачам сталого функціонування цих доволі складних і важливих для людства систем.

Теоретичний і практичний досвід традиційних методів прогнозування розвивається на базі новітніх інформаційних технологій математичного та системного моделювання. Особлива увага приділяється фундаментальним концепціям і принципам управлінського фінансування в тих галузях і структурах суспільства, в яких проблеми управління та прийняття рішень по функціонуванню починають ґрунтуватися на принципах економічної та екологічної значущості.

Завдяки прогресу в аналітичному моделюванні та росту обчислювальних можливостей інформаційних технологій сьогодні можна виконувати ефективні прогнози динамічних процесів, що нас цікавлять. Зокрема, застосовуючи різноманітні підходи до моделювання та прогнозування реальних процесів представлених часовими рядами, з різним рівнем впливовості параметрів на процес та складній структурі взаємозв'язків.

Будуючи різноманітні моделі прогнозування, можна враховувати та передбачувати аномальні вибухи в процесах, а також робити певні висновки щодо становища держави на прикладі економічних показників таких, як валовий внутрішній продукт, валовий національний продукт, індекс споживчих цін та інші, а для дослідження екологічної ситуації рівень забруднення повітря, води та рівень життя екосистеми. Кожен із показників

та характер його поведінки може мати свої особливості, які відображатимуться на загальному часовому ряді, відповідно й робота з такими даними буде відрізнятися та ускладнюватися. Для коректного моделювання та прогнозування існують критерії адекватності моделі та оцінки якості прогнозу, за допомогою цих показників можна оцінити рівень достовірності прогнозу та порівняти які побудовані моделі найкраще відображають характер поведінки даних.

РОЗДІЛ 1

ОСОБЛИВОСТІ МОДЕЛЮВАННЯ ЧАСОВИХ РЯДІВ ТА СУЧАСНІ МЕТОДИ ЇХ ПРОГНОЗУВАННЯ

1.1 Актуальність методів прогнозування в умовах сучасності

Сьогодні в умовах різкої зміни тенденцій життя в усіх сферах суспільства дуже важливе місце в управлінні будь-якої одиниці керівної ланки займає коректний аналіз та регулювання ресурсів різної сфери застосування. Адже успіх діяльності, окремо взятої інстанції великого механізму, напряду залежить від застосування нових принципів управління. Так, у зв'язку з появою непрогнозованих процесів у світі, останнім часом спостерігається неоднозначна динаміка екологічних та економічних процесів, що безпосередньо зазвичай негативно впливає на людську діяльність. Керівництво держав повинно звернути увагу на катаклізми в таких областях як екологія та економіка.

Якби керівні органи вели політику планування на випередження батьох світових проблем можна було запобігти, зокрема такі дії на запобігання проблем можливі завдяки якісному аналізу процесів, їх моделюванню з математичної точки зору та безпосередньо прогнозуванню.

Крім одержання можливих майбутніх оцінок тих чи інших досліджуваних параметрів, метою прогнозування також є спонукання до роздумів про те, що може статися у зовнішньому і внутрішньому середовищі, і до яких наслідків для економіки та екології це призведе. Прогнозування підвищує пильність системних агентів і, отже, їх здатність реагувати на зміни. Цей ефект досягається навіть тоді, коли план не виконаний у зв'язку з тим, що деякі гіпотези, покладені в основу прогнозного сценарію, що не реалізувалися.

Проблема прогнозування процесів внаслідок швидких, часом погано передбачуваних змін зовнішнього середовища, в останні роки стала особливо складної. З урахуванням цих труднощів і критичності помилок у прогнозах деякі фахівці були змушені заговорити про даремність прогнозування нестационарних процесів і складності структурних характеристик систем економіки та екології.

Проблемам аналізу і прогнозування нестационарних процесів в літературі, саме в контексті економічних та екологічних наук, приділено недостатню увагу. Кілька публікацій присвячено загальним теоретичним питанням про роль цих інструментів. Ведуться дискусії про переваги систематизації аналізу і прогнозування, але єдиного апарату дослідження нестационарних - еволюційних процесів поки не існує. Залишається неосвітленим ряд актуальних питань, наприклад, способи визначення характеристик процесу, механізм вивчення структури об'єкта, аналізу факторів, вибір методів прогнозування, підвищення ефективності прогнозів. Все, що було перелічено вище, визначає актуальність дисертаційного дослідження.

1.1.1 Проблеми екологічних процесів у світі

Наша планета вже існує понад 7 млрд років і протягом всього цього періоду дуже складно злічити всі глобальні проблеми, які зустрічались на її шляху. Проте нам досить легко вказати на ті проблеми, що безпосередньо стосуються нашого покоління, адже вони як ніколи є видимими, і причиною їхньої появи більшою мірою є саме людство. Проблеми екологічного характеру, пов'язані з діяльністю людини в глобальних масштабах, виникли не так давно і ми маємо можливість їх дослідити, проаналізувати та зробити

необхідні висновки, адже це безпосередньо стосується нашої теперішньої та майбутньої життєдіяльності. Екологічні катаклізми є наслідком виробничої діяльності людства як малого так і великого обсягу, що в деякій мірі з'явилися з розвитком технологій масового виробництва.

За оцінками вчених, людство в даний час живе за рахунок майбутніх поколінь, яким підготовлено набагато гірші умови життя, що, неминуче вплине на стан їх здоров'я і соціальне благополуччя. Щоб уникнути цього, людям потрібно навчитися існувати тільки на частину долі з основного капіталу - природи, при цьому не витрачаючи самий капітал.

Починаючи з XX століття, цей капітал витрачається неухильно зростаючими темпами, і до теперішнього часу природа Землі змінена настільки, що ось уже кілька десятиліть на міжнародному рівні обговорюються глобальні екологічні проблеми. У екосистемі, що ми використовуємо, навіть новітні технології раціонального природокористування не дозволяють зберегти біорізноманіття. Для цієї мети необхідні особливо охоронні природні території (ООПТ), в яких господарська діяльність повністю заборонена або обмежена.

Рівень впливу людини на навколишнє середовище залежить в першу чергу від технічної озброєності суспільства. Вона була вкрай мала на початкових етапах розвитку людства. Однак з розвитком суспільства, зростанням його продуктивних сил ситуація починає змінюватися кардинальним чином. XX століття - це століття науково-технічного прогресу. Пов'язане з якісно новим стосунками науки, техніки і технології, воно колосально збільшує можливі і реальні масштаби впливу суспільства на природу, ставить перед людством цілий ряд нових, надзвичайно гострих проблем, в першу чергу - екологічну.

Людині як організму життєвонеобхідними чинниками для існування є повітря та вода. Але як організм може нормально функціонувати, якщо кожен із них має високий рівень забрудненості і відповідно не виконує

належних функцій. Розглянемо більш детально глобальну екологічну проблему людства в розрізі цих двох життєвонеобхідних чинників.

Забруднення повітря - це проблема охорони здоров'я та навколишнього середовища у всіх країнах світу, але з великою різницею в рівні серйозності та відповідальності за наслідки.

Забруднення повітря - поєднання зовнішніх і внутрішніх твердих частинок та озону - є фактором ризику для багатьох провідних причин смерті, включаючи хвороби серця, інсульт, інфекції нижніх дихальних шляхів, рак легенів, діабет та хронічну обструктивну хворобу легень. Забруднення повітря спричиняє 5 мільйонів смертей щороку.

У країнах з низьким рівнем доходу ця проблема очолює список. У 2017 році воно спричинило 5 мільйонів смертей у всьому світі. Це означає, що це призвело до 9% - майже 1 із 10 - смертей. Рівень смертності від забруднення повітря найвищий у країнах з низьким та середнім рівнем доходу. Наслідки цієї проблеми сприяють 9% смертей у всьому світі - це варіюється від 2% до 15% у різних країнах. [1]

Забруднення повітря не лише зменшує тривалість життя людей, але і сильно впливає на якість життєдіяльності, адже низький рівень вмісту кисню в повітрі або ж його забрудненість токсичними газами значної мірою впливають на процеси функціонування організму та його органів.

На Рисунку 1.1 можна побачити показники смертності від забруднення повітря у всьому світі, що вимірюється як кількість смертей на 100 000 жителів певної країни чи регіону.

Death rates from air pollution

Death rates are measured as the number of deaths per 100,000 population from both outdoor and indoor air pollution. Rates are age-standardized, meaning they assume a constant age structure of the population to allow for comparisons between countries and over time.

Our World
in Data

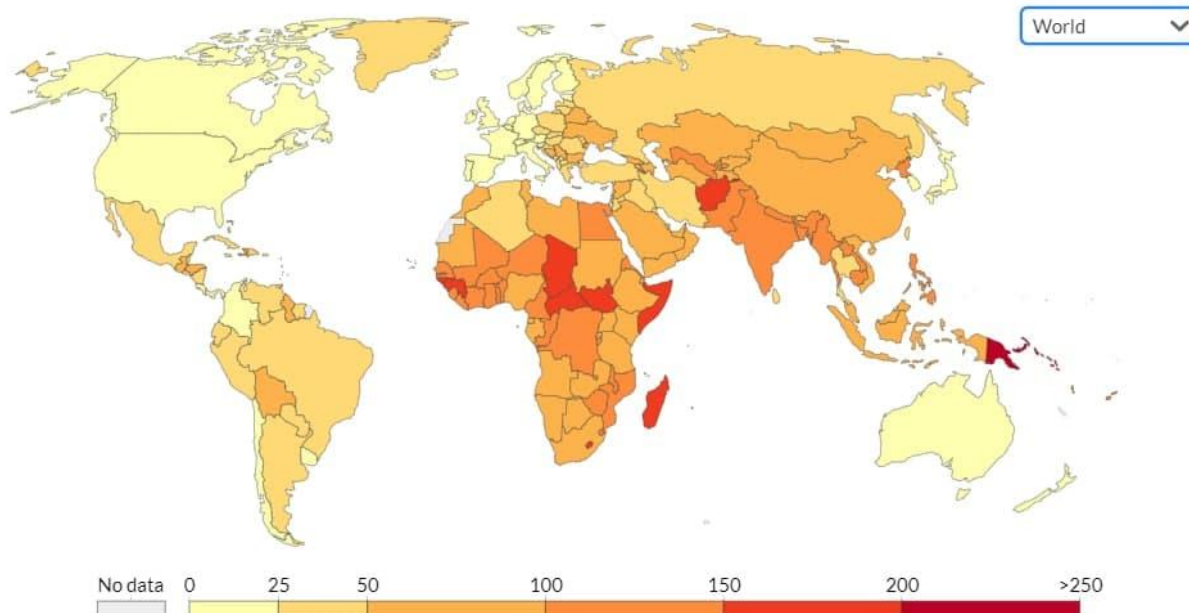


Рисунок 1.1 – Показники смертності від забруднення повітря у світі

Ми бачимо, що рівень смертності, як правило, найвищий у Африці на південь від Сахари та Південній Азії. Це підкреслює великі відмінності у всьому світі: рівень смертності в країнах із найбільшим показником перевищує показники забрудненості більшості країн Європи та Північної Америки у 100 разів.

Навантаження на забруднення повітря, як правило, стає більшим як у країнах з низьким, так і середнім рівнем доходу з двох причин: рівень забруднення приміщень, як правило, високий у країнах з низьким рівнем доходу через залежність від твердого палива для приготування їжі та обігріву оселі; а забруднення атмосферного повітря, як правило, зростає в міру індустріального розвитку країн та переходу від низьких до середніх рівнів доходів.

Вода непридатна до споживання є однією з найбільших у світі проблем охорони здоров'я та довкілля - особливо для країн з низьким рівнем життя.

Відсутність доступу до безпечних джерел води є провідним фактором ризику для інфекційних захворювань, включаючи холеру, діарею, дизентерію, гепатит А, черевний тиф та поліомієліт. У країнах з низьким рівнем доходів небезпечні джерела води становлять 6% смертей.

За оцінками, в 2017 році в результаті небезпечних джерел питної води загинуло 1,2 мільйона людей. Це становило 2,2% смертей у світі. Порівнюючи частку смертей, пов'язаних з небезпечною водою, з часом чи між країнами, ми порівнюємо не лише ступінь доступу до води, але її тяжкість у контексті інших факторів ризику смерті. Частка чистої води залежить не тільки від того, скільки передчасно помирає від неї, але від чого ще люди вмирають і як це змінюється.

Що стає зрозумілим, так це великі різниці у рівні смертності між країнами: показники високі в країнах з низьким рівнем доходу, особливо в Африці на південь від Сахари та Азії. Частота тут часто перевищує 50 смертей на 100 000 - у Центральноафриканській Республіці та Чаді це було більше 100 на 100 000.

Порівняйте це із показником смертності в країнах з високим рівнем доходу: у Європі показники смертності нижче 0,1 смертності на 100 000. Це більше ніж 1000-кратна різниця.

Яка частка людей має доступ до безпечної питної води? За підрахунками, лише 71% населення світу має доступ до безпечної питної води. Це означає, що 29% країн світу не мають доступу. [1] Доступ до вдосконаленого джерела води збільшує ймовірність того, що питна вода є чистою та безпечною. Але це не гарантує, що вона вільна від забруднення.

На Рисунок 1.2 ми бачимо частку людей у всьому світі, які мають доступ до безпечної питної води.

Share of the population using safely managed drinking water, 2015

A safely managed drinking water service is defined as one located on premises, available when needed and free from contamination.

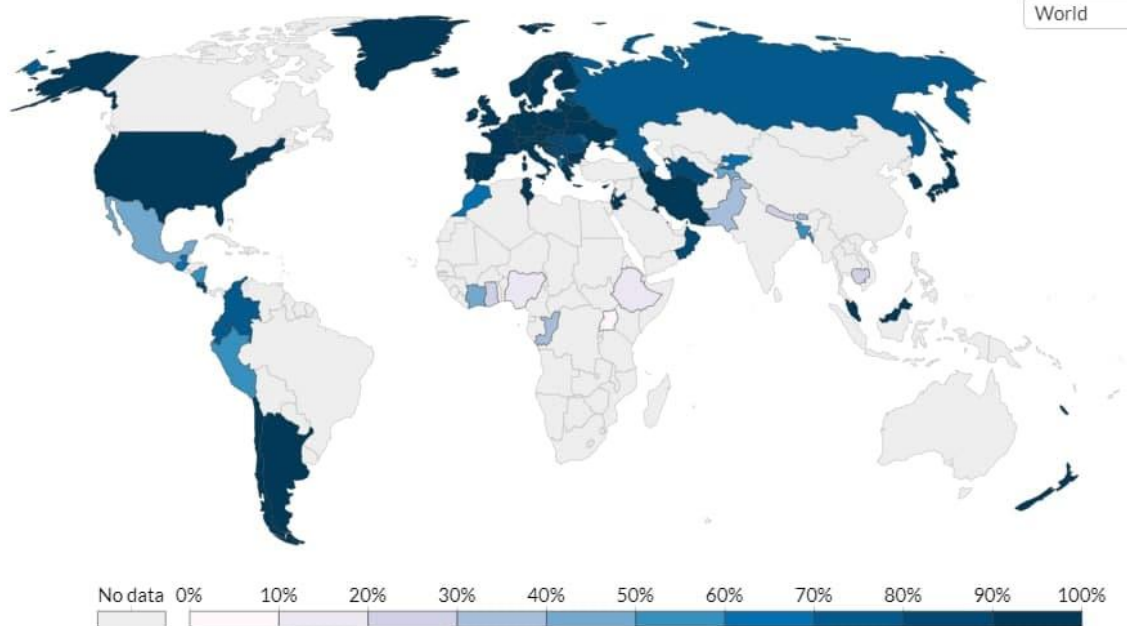


Рисунок 1.2 – Розподілення безпечної питної води

Доступ до питної води з покращеного джерела не гарантує, що вода є безпечною або достатньо безпечною, оскільки ці характеристики не перевіряються під час обстеження. Але поліпшені технології питної води частіше, ніж ті, що характеризуються як не покращені, забезпечують безпечну питну воду та запобігають контакту з людськими екскретами. Навіть у країнах з високим рівнем доходу не завжди можна пити очищену воду. Доступ до вдосконаленого джерела води прирівнюється до підключення до системи постачання; він не враховує відмінності у якості та вартості.

1.1.2 Проблеми економічних процесів у світі

Основною проблемою економічних процесів в державі, а також загалом в світі є те, що керівництво не може правильно проаналізувати та ефективно спланувати план розвитку держави, що призводить до бідності та незадовільного рівня розвитку населення та країни загалом. Планування повинно проводитись не окремо обраною владою на обмежений термін, а стати системою стратегічного розвитку на багато років вперед.

Для цього держави з високорозвиненою економікою створюють системи прогнозів, які розробляються для формування стратегії і тактики розвитку економіки на загальнодержавному рівні. Вона включає різні прогнози в тимчасовому аспекті (коротко-, середньо і довгострокові) і за рівнями управління (макроекономічні, міжгалузеві, галузеві, міжрегіональні, регіональні, на рівні підприємств), а також приватні (демографічні, прогнози природних ресурсів, екологічні прогнози, прогнози розвитку науки і техніки, соціальні, економічні прогнози) і комплексні прогнози. На макрорівні розробляються плани, програми економічного і соціального розвитку країни, на регіональному рівні - плани розвитку регіонів (областей, районів, міст). Розвиток окремих галузей відображується в галузевих планах, на мікрорівні розробляються плани розвитку підприємств. [2]

Економічне зростання. Фактори і проблеми темпів економічного зростання. Під зростанням розуміється зміна умов розвитку економіки, тенденції в розвитку окремих її елементів (цін, виробництва, безробіття та ін.). Економічне зростання являє собою зміну результатів функціонування економіки і використання ресурсів: збільшення валового внутрішнього продукту, національного доходу, доходу на душу населення, рівня життя. кінцевою метою економічного зростання є зростання споживання і добробуту населення.

У масштабі господарства країни в цілому економічне зростання характеризує збільшення обсягу виробництва товарів і послуг. Для ілюстрації цього процесу використовують поняття довгострокового тренда економічного розвитку. Трендова динаміка реального ВВП - це рівень, на якому б знаходився фактичний ВВП, якби він протягом всього періоду рівномірно зростав темпами, рівними рівнем середнього фактичного темпу приросту. [2]

Основними цілями соціально-економічного розвитку є досягнення стійкого економічного зростання, високого ступеня зайнятості, стабілізації цін, зовнішньоекономічної рівноваги і на цій основі - підвищення рівня життя народу. Система макроекономічних показників включає показники, що характеризують виробництво, зайнятість, рівень цін і заробітної плати, зовнішньоторговельні потоки, податкові надходження і державні витрати, грошову масу і ін. Макроекономічні показники служать для передбачення розвитку економіки, виявлення тенденцій, причин їх зміни і вироблення економічної політики з метою досягнення намічених цілей. Центральне місце в системі макроекономічних показників відводиться показниками, які характеризують розвиток виробництва продукції і послуг.

Економічний добробут вимірюється як валовий внутрішній продукт (ВВП) на душу населення, вартість усіх товарів і послуг, вироблених країною за один рік, поділена на населення країни. Економічне зростання - міра зміни ВВП з року в рік. Цей запис показує, що сучасний досвід економічного зростання є абсолютним винятком у дуже довгостроковій перспективі соціальної історії.

Нижче рисунку 1.3 представлені оцінки економічного процвітання в довгостроковій перспективі, які були опубліковані Базою даних проекту Меддісон у своєму великому оновленні в 2018 році.

GDP per capita, 2016

GDP per capita adjusted for price changes over time (inflation) and price differences between countries – it is measured in international-\$ in 2011 prices.

Our World
in Data

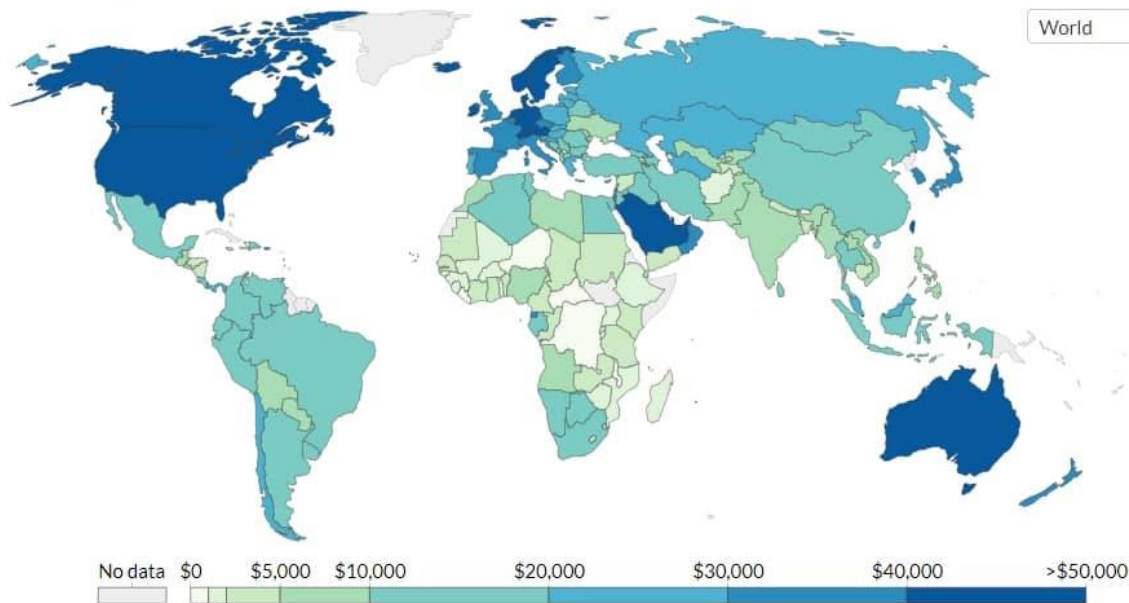


Рисунок 1.3 – ВВП на душу населення з урахуванням інфляції в міжнародних доларах

У 1870 р. світовий ВВП на душу населення, за оцінками, становить близько 1263 доларів на рік, і це вже після того, як деякі регіони світу досягли економічного зростання. За всі сотні, а в дійсності тисячі років до 1870 року середній ВВП на душу населення був ще нижчим.

Процвітання - це зовсім недавнє досягнення, яке відрізняє останні 10 або 20 поколінь від усіх їхніх предків. У 2016 році середній ВВП на душу населення становив 14574 доларів - більше, ніж у 10 разів перевищував середній показник минулого. Дохід середньостатистичної людини у світі зріс із лише 3300 доларів в 1950 році до 14 574 доларів США у 2016 році. Що означає, що середньостатистична особа в 4.4 рази багатша, ніж більше як пів століття тому.

Якщо проаналізувати ВВП на душу населення в Англії, то можна побачити, що на початку XIV століття рівень доходів мав значний скачок. Доходи зросли приблизно на третину лише за кілька років, це було спричинено чумою. За три роки після 1348 року чисельність населення

зменшилася з 8 до 4,3 мільйонів. Тобто динамічні зміни в економічних процесах значною мірою залежать від глобальних катаклізмів, пов'язаних із народонаселенням.

ВВП на душу населення в США напередодні незалежності все ще був нижче 2000 доларів, скоригований на інфляцію та виміряний у цінах 2011 року, оцінюється у 1883 долари. У 2016 році - 240 років після здобуття незалежності - ВВП на душу населення збільшився більш ніж у 28 разів - до 53 015 доларів. Це означає, що дохід на одну людину за один рік у минулому був меншим, ніж результат середньостатистичної людини за два тижні сьогодні. Примітно, наскільки стабільним було економічне зростання протягом цього дуже тривалого періоду. З 1870 по 2016 рік ВВП на людину в економіці США зростав в середньому на 1,83 відсотка на рік, лише з дуже короткими відхиленнями від цієї дуже стійкої тенденції.

Економічне зростання має бути досягнуте в той час, коли нам необхідно терміново зменшити свій вплив на навколишнє середовище. Це означає, що важливим є не лише темпи зростання. Як каже Маріана Мацзукато, "економічне зростання має не лише темпи, але й напрямок". І багато шляхів зростання вказують у напрямку, який не збільшує наш екологічний збиток, а натомість часто може зменшити вплив (кращий догляд за хворими та літніми, кращі навчальні заклади, турбота про психічне здоров'я, вдосконалення сонячних технологій).

Ця статистика демонструє нам, що економічний ріст дуже різниться між країнами та необхідно зважати хто з часом отримує вигоду від економічного зростання. У різних країнах присутня велика залежність економічного зростання та перерозподілу цих прибутків між усіма прошарками суспільства. Ця нерівномірна залежність неабиякий раз доводить, що планування росту економіки та правильне керування необхідне для розсудливого управління державою, адже воно значної мірою може покращити становище населення та їх благоустрій.

1.2 Існуючі методи прогнозування та особливості економічних та екологічних процесів

Прогнозування економічних та екологічних процесів в основному зводиться до аналізу та дослідження параметрів показників, що характеризують ці процеси. Вони можуть бути найрізноманітніші і впливати на прогноз як окремий показник так і сукупність взаємозв'язків. Основні з яких це: економічний ріст, національний дохід (дохід громадян за певний проміжок часу), ВВП (валовий національний продукт), ВВП (валовий внутрішній продукт), індекс конкурентноспроможності, рівень безробіття – для економіки; рівень забрудненості повітря (з багатьма побічними складовими хімічних сполук), рівень забрудненості води (як питної та промислової), потужність земельних ресурсів та інші – для екології.

Оскільки за допомогою прогнозування даних показників можна ефективніше керувати економічною та екологічною ситуацією як в державі так і в інших структурах, результати такого прогнозування мають велике значення в державотворенні та нормальному функціонуванні об'єктів, які в ньому задіяні.

Для прогнозування природних процесів, що впливають на стан екології середовища використовують наступні методи [3]:

- метод аналогів;
- метод екстраполяції;
- метод експертних оцінок;
- картографічний метод;
- математичний метод;
- просторово-часовий прогноз з використанням геоінформаційних систем.

Метод аналогів використовується в зв'язку з можливістю перенесення закономірностей справедливих для просторових рядів, на тимчасові ряди і навпаки, виходячи з близькості просторового і тимчасового прогнозу; ступінь подібності аналога і об'єкта прогнозу визначається однорідністю порівнюваних об'єктів. Так для Ангарський водосховищ як аналог прийнято водосховище Братської ГЕС. Критерієм подібності може бути режим рівнів водосховищ, геологічні структури, таксаційні характеристики зон затоплення і зон, схильних до впливу ветроволнових явищ, клімат, схожість антропогенних навантажень і т.п. Основою для успішного застосування цього методу є достовірність даних, взятих для аналога.[4]

Метод експертних оцінок заснований на зборі інформації групової оцінки події. Однак теоретичного обґрунтування точності і надійності методів групової оцінки не існує. Єдиним критерієм є експериментальна перевірка за характерними басейнах. Метод використовувався при оцінці запасів затонулої деревини, виведеним з молевого сплаву.

Метод екстраполяції передбачає поширення сформованих в часі закономірностей в передпрогнозний період на певний період в майбутньому. При екстраполяційном методі прогнозування визначальним є вибір головних прогнозованих факторів.

Картографічний метод - найбільш розповсюджений комплексний метод, частина кожного вищевказаного методу прогнозування. Цей метод використовується для загальної оцінки просторового розвитку процесу.

Метод математичного моделювання заснований на доцільному абстрагуванні процесів розвитку в майбутньому. У практиці управління водними ресурсами, водним господарством, водокористуванням математичне моделювання використовується при плануванні, проектуванні, експлуатації водогосподарських систем, прогнозуванні водокористування, наслідків здійснення водогосподарських заходів і при вирішенні інших завдань. У наші дні для екологічного прогнозування відкриваються можливості використання

сучасних географічних інформаційних систем (ГІС), що дозволяють виконувати просторово-тимчасовий прогноз.

Пошуковий прогноз – метод ідентифікації усіх можливих станів будь-якого економічного середовища в майбутньому. У цьому понятті закладено передбачення досить умовного продовження майбутніх тенденцій перебігу процесів, які досліджуються в минулому або ж в реальному часі. Але досить важливим є здатність до абстракції по відношенню до рішень та дії, за допомогою яких можлива радикальна зміна трендів. У такому випадку розглянутий прогноз буде інтерпретувати як розвиватиметься процес з найбільшою вірогідністю за умови збереження трендів, що вже існують. Також ці прогнози мають іншу назву – варіативні, тобто сценарні розрахунки.[3]

Нормативний прогноз – метод ідентифікації шляхів та розрахунку часу досягнення усіх можливих станів. Основною метою даного виду прогнозу є знайти шляхи для досягнення бажаного результату.

Структура пошукового прогнозу має вигляд певної градації можливостей, після побудови якої вже формуються різні рівні ймовірностей середовища або явища, що прогнозується. Коли формується нормативне прогнозування, то в ньому формування рівнів ймовірностей розпочинається з постановки конкретного стану й переходу до спостереження трендів.

Розглянуті види прогнозів ефективно застосовуються на практиці одночасно, доповнюючи одне одного, але не як інструмент прогнозування, а скоріше як підхід до створення прогнозу.

За періодом прогнозування - інтервалу часу, тобто на скільки кроків вперед буде розрахований прогноз досліджуваного явища, виділяють оперативні (поточні), короткострокові, середньострокові, довгострокові прогнози.

Розрахування оперативних прогнозів відбувається на перспективу, тобто передбачається, що протягом вказаного періоду часу досліджуваний

об'єкт не буде мати істотних змін в динаміці. Це стосується як кількісних показників, так і якісних.

У свою чергу короткострокові прогнози розглядаються на перспективу тільки кількісних змін динаміки об'єкта, а довгострокові переважно якісних змін, проте наявні і кількісні. Середньострокові прогнози є чимось проміжним між короткостроковими та довгостроковими, але вони мають перевагу в значимості кількісних змін над якісними. Проте, коли у середньостроковому прогнозі очікуються значні якісні зміни, то його можна розглядати в перспективі, як довгостроковий, але в загальних рисах розвитку об'єкту.

Для виконання соціально-економічного прогнозування за допомогою досліджень було встановлено наступні часові по кожному з видів прогнозу: оперативні прогнози - менше місяця, короткострокові - менше року, середньострокові – на декілька (зазвичай менше п'яти) років, довгострокові – від п'яти до 15-20 років, далекострочковий - більше 20 років.

Для прогнозування державних процесів, аналітики та державні структури різного ступеню управління, найчастіше застосовують короткострокові прогнози, які є характерними для країн з перехідною економікою. Що стосується виконання прогнозу на термін функціонування конкретно діючої ланки правління або ж його державного органу, то в такому випадки використовуються середньострокові прогнози. Довгострокове прогнозування ефективно використовують для дослідження демографічних процесів, а саме народонаселення, рівня народжуваності та смертності на десятки років вперед.[5]

За об'єктом дослідження виділяють іноваційні та соціальні види прогнозів. Соціальним прогнозам притаманні характеристики, які можуть безпосередньо впливати на реалізацію цих прогнозів, тобто вони можуть здійснюватись без стороннього втручання або ж навпаки під впливом

людських дій, а саме прагнення виконання плану або досягнення певної мети руйнуватися. Для цього формуються спеціальні плани та програми.

За масштабом прогнозування виділяють: макро-економічні (національної економіки) та структурні (міжгалузеві, міжсекторальні, міжрегіональні) прогнози, прогнози розвитку окремих комплексів, секторів і регіонів, прогнози господарюючих суб'єктів, а також окремих виробництв і продуктів. Відзначимо, що об'єкти макроекономіки більш стійкі і інерційні в своєму розвитку в порівнянні з об'єктами мікроекономіки.

Розподіл прогнозів залежно від характеру об'єктів пов'язаний з різними аспектами відтворювального процесу. За цією ознакою виділяють такі економічні прогнози: розвитку виробничих відносин, соціально-економічних передумов і наслідків технологічного прогресу, динаміки економічної системи (темпів, факторів і структури), відтворення трудових ресурсів, зайнятості, економічного використання природних ресурсів, інвестицій, рівня життя населення, доходів і цін, споживчого попиту, зовнішньоекономічних зв'язків та інші. Слід мати на увазі при цьому, що різниця та незалежний розгляд окремих елементів системи неспроможні з методологічної точки зору.

1.3 Огляд деяких комп'ютерних систем для побудови моделей прогнозування та аналізу даних

Оскільки питання аналізу, моделювання та прогнозування різноманітних процесів у світі є досить актуальне, то різноманітні аналітичні компанії намагаються створювати найзручніше програмне забезпечення для вирішення даних питань. Але проблема універсальності системи, а також можливості для обробки великих масивів даних та подальшою роботою з

ними, залишилася. Деякі із компаній намагаються вдосконалити свою апаратну частину та створити максимально інтуїтивно зрозумілий інтерфейс та функціонал для зручної роботи з даними як для людей, які займаються в цій сфері тривалий час (аналітики, науковці, працівники відповідних установ) та для початківців (студенти, викладачі, комерційні підприємства). Досить непогані результати мають наступні програми – EViews та Statistica.

EViews підтримує широкий спектр базових статистичних аналізів, що охоплює все: від простої описової статистики до параметричних та непараметричних тестів та гіпотез. Основна описова статистика швидко та легко обчислюється для цільової вибірки шляхом класифікації на основі однієї або декількох змінних, перетину або періоду в панельних або об'єднаних даних. Тести на гіпотези щодо середнього, медіани та дисперсії можуть проводитись, включаючи тестування на конкретні значення, тестування на рівність між рядами або тестування на рівність в межах однієї серії при класифікації за іншими змінними (що дозволяє виконати односторонню ANOVA). Інструменти для коваріації та факторного аналізу дозволяють вивчити взаємозв'язки між змінними.

EViews допоможе візуалізувати розподіл даних, використовуючи гістограми, теоретичний розподіл, щільність ядра або кумулятивний розподіл, графік виживання та квантілі. QQ-графіки (квантильно-квантильні графіки) можуть бути використані для порівняння розподілу пари серій або розподілу однієї серії проти різноманітних теоретичних розподілів.

За допомогою EViews можна досліджувати властивості часових рядів за допомогою інструментів, починаючи від простих графіків автокореляції, закінчуючи частотними фільтрами, Q-статистикою та модульними тестами кореневої системи.

EViews надає функції автокореляції та часткової автокореляції, Q-статистику та функції перехресної кореляції, а також модульні кореневі тести (ADF, Phillips-Perron, KPSS, DFGLS, ERS або Ng-Perron для одиничних

часових рядів та Levin-Lin-Chu, Breitung, Im-Pesaran-Shin, Fisher, або Hadri для панельних даних), тести на коінтеграцію (Йохансен з критичними значеннями Маккіннона-Хауга-Мікеліса та р-значеннями для звичайних даних, а Педроні, Као або Фішер для панельних даних) , причинно-наслідкові зв'язки та тести незалежності. [6]

Якщо дисперсія вашої серії коливається з часом, EViews може оцінити шлях дисперсії, використовуючи широкий спектр моделей авторегресивної умовної гетероскедастичності (ARCH). EViews обробляє специфікації GARCH (p, q), EGARCH (p, q), TARARCH (p, q), PARARCH (p, q) та компонент GARCH та забезпечує максимальну оцінку ймовірності помилок, що виникають після нормальної, t або генералізованої помилки поширення. Загалом рівняння моделей ARCH може включати терміни ARCH та ARMA, і як середнє, так і рівняння дисперсії враховують екзогенні змінні. STL Decomposition забезпечує сезонне коригування для будь-яких частотних даних, а просте сезонне коригування з використанням адитивних та мультиплікативних різницевих методів також підтримується в EViews.

EViews можна використовувати для обчислення тенденцій та циклів за даними часових рядів, використовуючи фільтри Ходріка-Прескотта, Бакстера-Кінга, Крістіано-Фіцджеральда з фіксованою довжиною та Крістіано-Фіцджеральда, асиметричні фільтри повного пропуску смуг (частоти).

В основі інноваційного інтерфейсу EViews лежить поняття об'єкта. Ряди, рівняння та системи - це лише кілька прикладів об'єктів. Кожен об'єкт має своє вікно, меню, процедури та власні уявлення про свої дані. Більшість статистичних процедур - це просто альтернативні уявлення про об'єкт. Наприклад, простий вибір меню з вікна серії змінює відображення між електронною таблицею, різними видами графіків, описовою статистикою та тестами, таблицями, корелограмами, коренем одиниці та тестами на незалежність.

Аналогічним чином, вікно рівняння дозволяє перемикатися між відображенням специфікації рівняння, основними результатами оцінки, фактично встановленими залишковими графіками та таблицями, відображенням структури рівняння ARMA (за необхідності), градієнтами та похідними специфікації, матриця коваріаційних коефіцієнтів, графіки прогнозів та оцінки, а також понад десяток діагностичних та гіпотезних тестів.

На відміну від традиційних програм статистики, які підтримують перегляд одночасно лише одного рівняння оцінки або графіка, EViews дозволяє одночасно відображати декілька об'єктів, кожен у своєму вікні. Ця справжня підтримка декількох вікон дозволяє легко виконувати паралельне порівняння серійних графіків, тестів гіпотез, оцінок рівнянь або прогнозів моделей, розроблених за альтернативних припущень. EViews включає найкращі сучасні електронні таблиці та технології реляційних баз даних в інструменти для виконання традиційних завдань статистичного програмного забезпечення. Завдяки Об'єктно-орієнтованому підходу EViews включає складну технологію взаємозв'язків, яка дозволяє визначити їх між кількома об'єктами та зовнішніми джерелами даних. Наприклад, об'єкти серій можуть бути пов'язані за формулою з даними інших серій, збігатися з об'єднаними або перетвореними даними з альтернативних наборів даних або з даними із зовнішніх баз даних. Завдяки чому усі задіяні дані оновлюються при зміні якогось з об'єктів. Також дана платформа має потужну інтеграцією Windows, включаючи імпорт файлів перетягування для понад двадцяти популярних форматів файлів, копіювання та вставлення експорту графіків та таблиць якості презентацій.

Statistica - це набір програмних продуктів та рішень для аналітики великих масивів даних, управління даними, інтуїтивно зрозумілою візуалізацією даних та процедур виокремлення даних, а також різноманітні

методи прогнозуючого моделювання, кластеризації, класифікації та дослідницьких методів.

За допомогою Statistica ви можете проводити спеціальний аналіз по групах практично з будь-якого діалогового вікна результатів, перегляд результатів для всіх груп, об'єднаних або розбитих за однією або кількома змінними групування. Ця дуже потужна функція для дослідницького аналізу даних може бути використана для перевірки узгодженості результатів між групами для будь-якого аналізу.

Statistica відмінно працює при аналізі дуже багатьох проблем з даними (наприклад, дуже великих, незбалансованих) експериментальні конструкції або складні аналізи, які вимагають повторення обчислень. Тому, як правило, корисно починати кожен дослідницький аналіз з обчислень: прості описові статистичні дані, частотні таблиці та графіки для розуміння структури даних і визначити кількість унікальних груп (поєднання значень у групування змінних) у даних.

Також Statistica надає доступ до шести способів обробки вихідних даних. Ви можете налаштувати спосіб управління вихідними даними в інтерфейсі:

- Workbooks;
- автономні вікна;
- звіти;
- Microsoft Word;
- веб-переглядач;
- система управління документами SharePoint або Statistica (SDMS).

З використанням Statistica можна обирати моделі для імплементації поведінки обраних даних. Наприклад, C&RT, Boosted tree та нейронна мережа за замовчуванням вибрані як моделі або алгоритми, які автоматично перевірятимуться.

Обчислення для побудови прогнозних моделей можна виконати або локально або на сервері Statistica Enterprise. Також можна перевірити результати моделювання за допомогою класифікаційних показників найкращої моделі на наборі даних перевірки.

Висновки до розділу 1 і постановка задачі

У даному розділі було досліджено та продемонстровано екологічну та економічну ситуацію у світі. Зокрема, представлено статистику залежності забруднення повітря та води із станом здоров'я людей окремих країн та вплив економічного росту, виражений у ВВП країни, на рівень життя населення. Представлено характерні особливості поведінки економічних та екологічних процесів та їхнє відображення на динаміці статистичних даних. Було доведено важливість питання якісного прогнозування даних процесів, а також коректний аналіз наслідків, в контексті ефективного управління на державному рівні. Розглянуто методи прогнозування, що характерні для природних процесів, та їх класифікація. Також було описано деякі комп'ютерні системи для побудови моделей прогнозування та аналізу даних, які сьогодні ефективно використовуються.

Постановка задачі

1. Виконати аналіз підходів і методів моделювання економічних та екологічних процесів в умовах нелінійності й нестационарності.
2. Знайти статистичні дані, які найкраще характеризують особливості розглянутих процесів та підготувати їх до моделювання.
3. Розглянути математичні методи моделювання та прогнозування часових рядів відповідно до особливостей обраних даних та

навести порівняльну характеристику, переваги та недоліки кожного з них.

4. Виконати практичне дослідження згідно етапів реалізації системи прогнозування.
5. Здійснити оцінювання якості прогнозування економічних та екологічних процесів за допомогою критеріїв оцінки прогнозу.
6. Продемонструвати результати роботи розглянутих методів у вигляді графіків та порівняльних таблиць на основі реальних статистичних даних.
7. Зробити висновки по дослідженню та підкреслити важливість якісного прогнозування економічних та екологічних процесів на рівні держави, виробивши поради стосовно подальших дій.

РОЗДІЛ 2

МЕТОДИ ТА ПІДХОДИ ДО МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ

2.1 Побудова авторгресійних моделей AR, ARMA, ARMAX

Розглянемо простий приклад рівняння AR(1):

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k), \quad E[\varepsilon(k)] = 0 \quad (2.1)$$

де $\{\varepsilon(k)\}$ – послідовність білого шуму з нульовим середнім.

Збільшимо незалежну змінну k , яка має зміст часу, на одиницю і запишемо рівняння знову:

$$y(k+1) = a_0 + a_1 y(k) + \varepsilon(k+1)$$

При відомих a_0, a_1 можемо знайти математичне сподівання $y(k+1)$ до дискретного моменту часу k .

$$E_k[y(k+1)] = a_0 + a_1 E_k[y(k)] = a_0 + a_1 y(k)$$

адже $y(k)$ в момент часу k є відомою константою.

Аналогічно з виразом (2.1) запишемо для моменту $k+2$:

$$y(k+2) = a_0 + a_1 y(k+1) + \varepsilon(k+2)$$

і з допомогою нескладних арифметичних дій знайдемо математичне сподівання $y(k+2)$:

$$E_k[y(k+2)] = a_0 + a_1 E_k[y(k+1)] = a_0 + a_1 E_k[a_0 + a_1 y(k)] = a_0 + a_0 a_1 + a_1^2 y(k)$$

Отже, тепер ми можемо записати вираз для загального випадку прогнозування на s кроків:

$$E_s[y(k+s)] = a_0 \left(\sum_{i=0}^{s-1} a_1^i \right) + a_1^s y(k) = a_0 \sum_{i=0}^{s-1} a_1^i + a_1^s y(k) \quad (2.2)$$

Вираз (2.2), що ми отримали, є функцією прогнозування на довільне число кроків. Прогнозування являє собою збіжний процес, якщо $|a_1| < 1$, отже:

$$\lim_{s \rightarrow \infty} E_k[y(k+s)] = \frac{a_0}{1-a_1} \quad (2.3)$$

де a_1 – знаменник геометричної прогресії.

Із отриманого виразу (2.3), можна зробити висновок, для будь-якого стаціонарного процесу АР або ж АРКС, оцінка умовного прогнозу асимптотично ($s \rightarrow \infty$) збігається до безумовного середнього. [16]

Доцільно обчислити похибку прогнозу за умові, що умовне математичне сподівання $E[\varepsilon(k)] = 0$:

$$f_k(s) = y(k+s) - E_k[y(k+s)].$$

Розрахуємо похибку прогнозу, де $s = 1$:

$$\begin{aligned} f_k(1) &= y(k+1) - E_k[y(k+1)] = a_0 + a_1 y(k) + \varepsilon(k+1) - a_0 - a_1 y(k) = \\ &= \varepsilon(k+1). \end{aligned}$$

Розрахуємо похибку прогнозу, де $s = 2$:

$$\begin{aligned} f_k(2) &= y(k+2) - E_k[y(k+2)] = \\ &= a_0 + a_1[a_0 + a_1y(k) + \varepsilon(k+1)] + \varepsilon(k+2) - E_k[y(k+2)] = a_0 + \\ &\quad + a_0a_1 + a_1^2y(k) + a_1\varepsilon(k+1) + \varepsilon(k+2) - a_0 - \\ &\quad - a_0a_1 - a_1^2y(k) = \varepsilon(k+2) + a_1\varepsilon(k+1). \end{aligned}$$

Отже, можна записати рівняння загального виду похибки прогнозу з довільною кількістю кроків прогнозування:

$$f_k(s) = \varepsilon(k+s) + a_1\varepsilon(k+s-1) + a_1^2\varepsilon(k+s-2) + \dots + a_1^{s-1}\varepsilon(k+1).$$

Тепер можна розрахувати дисперсію похибки прогнозування, врахувавши математичне сподівання $E_k[f_k(s)] = 0$. Отже оцінка прогнозу, яка обчислюється за виразом (2.2), є незміщеною. Маємо наступний вираз дисперсії похибки:

$$\text{Var}[f_k(s)] = \sigma^2[1 + a_1^2 + a_1^4 + a_1^6 + \dots + a_1^{2(s-1)}],$$

тобто дисперсія є функцією s . Асимптотичне значення дисперсії похибки прогнозу для стаціонарного процесу:

$$\lim_{s \rightarrow \infty} \text{Var}[f_k(s)] = \frac{\sigma^2}{1 - a_1^2}$$

де a_1^2 – знаменник геометричної прогресії.

Ковзне середнє використовується в економіці та фінансах для згладжування цінових рядів, короткострокових коливань, акцентуючи увагу на основних тенденціях і циклах. При обчисленні ковзного середнього значення функції розраховується кожен раз заново, враховуючи при цьому

кінцеву множину попередніх значень. Ковзне середнє певним чином «рухається» по часовому ряду. Оскільки при розрахунку береться середнє значення набору величин, то воно певним чином перебуває в постійному русі відносно часу. У процесах де закономірність динамічних рухів мало помітна, ковзне середнє змінюється у горизонтальному інтервалі.

Загальний вигляд формули для зваженого ковзного середнього:

$$MA(k) = \frac{\sum_{i=1}^N w_i \cdot y(k - i + 1)}{\sum_{i=1}^N w_i}$$

де N – розмір вікна ковзного середнього;

w_i – вагові коефіцієнти;

y – часовий ряд вхідних даних.

Побудуємо функцію прогнозування моделі АРКС(2,1):

$$y(k) = a_0 + a_1 y(k - 1) + a_2 y(k - 2) + \varepsilon(k) + \beta_1 \varepsilon(k - 1),$$

де $\varepsilon(k)$ – білий шум з нульовим середнім;

$y(0) = y_0$ – відома початкова умова.

Збільшимо незалежну змінну k , яка має зміст часу, на одиницю і запишемо рівняння знову:

$$y(k + 1) = a_0 + a_1 y(k) + a_2 y(k - 1) + \varepsilon(k + 1) + \beta_1 \varepsilon(k)$$

Тепер можемо знайти математичне сподівання $y(k + 1)$ до дискретного моменту часу k , рівняння якого матиме вигляд:

$$E_k[y(k + 1)] = a_0 + a_1 y(k) + a_2 y(k - 1) + \beta_1 \varepsilon(k),$$

де $\varepsilon(k)$ відома величина на момент часу k .

$$E_k[\varepsilon(k+j)] = 0, \forall j > 0$$

Запишемо рівняння для $k+2$:

$$y(k+2) = a_0 + a_1 y(k+1) + a_2 y(k) + \varepsilon(k+2) + \beta_1 \varepsilon(k+1)$$

Математичне сподівання для $k+2$:

$$\begin{aligned} E_k[y(k+2)] &= a_0 + a_1 E_k[y(k+1)] + a_2 E_k[y(k)] = \\ &= a_0 + a_1 [a_0 + a_1 y(k) + a_2 y(k-1) + \beta_1 \varepsilon(k)] + a_2 y(k) = \\ &= a_0 + a_0 a_1 + a_1^2 y(k) + a_1 a_2 y(k-1) + a_1 \beta_1 \varepsilon(k) + a_2 y(k) = \\ &= a_0(1 + a_1) + (a_1^2 + a_2) y(k) + a_1 a_2 y(k-1) + a_1 \beta_1 \varepsilon(k) \end{aligned}$$

Знайдемо математичне сподівання на три кроки вперед:

$$\begin{aligned} E_k[y(k+3)] &= a_0 + a_1 E_k[y(k+2)] + a_2 E_k[y(k+1)] = \\ &= a_0(1 + a_1 + a_1^2 + a_2) + (a_1^3 + 2a_1 a_2) y(k) + \\ &\quad + (a_1^2 a_2 + a_2^2) y(k-1) + \beta_1 (a_1^2 + a_2) \varepsilon(k). \end{aligned}$$

Отримавши попередні формули умовних математичних сподівань і провівши аналіз, можна вивести формулу для загального умовного математичного сподівання на s кроків, яка матиме наступний вигляд [16]:

$$E_k[y(k+s)] = a_0 + a_1 E_k[y(k+s-1)] + a_2 E_k[y(k+s-2)] \quad (2.4)$$

Якщо корені отриманого характеристичного рівняння (2.4) знаходяться всередині одиничного кола, то оцінка прогнозу асимптотично збігається до безумовного середнього значення

$$\lim_{s \rightarrow \infty} E_k[y(k+s)] = \frac{a_0}{1-a_1-a_2},$$

а для довільного процесу АРКС (p,q) оцінку умовного прогнозу можна записати як:

$$E_k[y(k+s)] = a_0 + \sum_{i=1}^p a_i E_k[y(k+s-i)]$$

Модель ARMAX(p,q,d) у загальному вигляді:

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + mv(k) + \sum_{j=1}^q b_j \cdot mv(k-j) + \sum_{s=1}^d c_s x_s, \quad (2.5)$$

де p – порядок авторегресійної частини;

q – порядок ковзного середнього;

d – кількість включених пояснюючих змінних.

ARMAX(AutoRegressive Moving Average model with eXogenous inputs model) модель, яка характерна наявністю екзогенного(зовнішнього) фактору.

У виразі (2.5) $\sum_{s=1}^d c_s x_s$ представляє собою лінійну комбінацію зовнішніх пояснюючих змінних x_1, \dots, x_d . У реальних демографічних процесах вони можуть являти собою додаткові чинники впливу на загальну картину демографії, такі як вплив ВВП країни, фінансування сфери медицини, покращення освіти, ріст кількості освічених людей в країні, роль жінки в суспільстві.

На основі аналізу сумісної кореляції вихідного сигналу y та x_s , що записується наступним чином $correl(y, x_s) = r_{y, x_s}$, було вирішено включити в модель ARMAX відповідну пояснювальну змінну x_s . Якщо $r_{y, x_s} > 0.5$, то змінну x_s необхідно включати.

У загальному випадку до складу ARMAX рівняння окрім регресорів x_1, \dots, x_d також можуть включатися лагові змінні $x_s(k - m)$. Для того, щоб включити відповідну авторегресійну частину регресора x_s необхідно виконати аналіз ЧКФ(y, x_s)[11].

2.2 Прогнозування на основі нейронних мереж

Складність математичних гіпотез про особливості часових рядів зазвичай зменшують сферу застосування класичних підходів та методів до прогнозування динамічних рядів. Застосування нейронних мереж в даній задачі обумовлено наявністю в більшості часових рядів складних закономірностей, які не виявляються відомими лінійними методами. Обробка інформації за допомогою методів нейронних мереж в широкому сенсі почала застосовуватись кілька десятиків років тому. Проте дослідницький інтерес до технологій штучних нейронних мереж мав не постійну основу, був час коли технології нейромереж розвивалися з великою інтенсивністю, а були часи, коли і взагалі інтерес зникав. Такий мінливий інтерес прямо пов'язаний з результатами імперичних досліджень. Сьогодні можливості інформаційних технологій, зокрема за допомогою нейронних мереж, застосовуються в більшості галузей науки: біологія, екологія, фінанси та економіка. За рахунок того, що технології нейромереж здатні до багатогранної обробки інформації, їм властива здатність до узагальнення та знаходження прихованих взаємозв'язів між вхідними та вихідними змінними.

Великою перевагою нейромереж є те, що вони мають можливість до навчання і узагальнення накопичених знань.

При вирішенні завдань прогнозування аналітику доводиться приймати рішення щодо таких характеристик часового ряду як тренд, сезонна і циклічна компоненти, робити припущення про модель часового ряду - адитивної, мультиплікативної та інші. Автоматичного способу виявлення трендів у тимчасових рядах не існує. У той же час при вивченні кривої, що відображає результати спостережень, аналітику важко робити припущення щодо повторюваності вигинів кривої через рівні проміжки часу. У прогнозуванні на основі нейронних мереж негативний вплив цих моментів намагаються мінімізувати.

2.2.1 Нейронні мережі. Рекурентна нейронна мережа

Нейронні мережі (NN) - це сукупність алгоритмів, які певною мірою відображають процеси людського мозку і використовуються для розпізнавання шаблонів. Вони імплементують сенсорні дані за допомогою машинного сприйняття, створення міток або кластеризації необроблених вхідних даних. Вони можуть розпізнавати числові шаблони, що містяться у векторах, в які повинні бути перекладені всі дані з реального світу (зображення, звук, текст або часові ряди). Штучні нейронні мережі складаються з великої кількості сильно взаємопов'язаних процесорних елементів (нейронів), що працюють разом, для того щоб вирішити проблему. Нейронна мережа, як правило, включає велику кількість процесорів, що працюють паралельно і розташовані по рівнях.

Штучна нейронна мережа - це математична модель (а також її програмна або апаратна імплементация) яка являє собою систему з'єднаних і

взаємодіючих між собою простих процесорів (штучних нейронів). Штучна нейронна мережа складається зі штучних нейронів та зв'язків між ними (синапсами). При цьому, кожному зв'язку відповідає певна величина w_i (вага зв'язку). Коли сигнал (числа, що надходять до нейрону) проходять по цьому зв'язку, його величина множиться на вагу зв'язку.

Існує багато типів нейронних мереж, які краще застосовувати для окремих типів задач, деякі з мереж наведено нижче:

- нейронні мережі прямого поширення;
- рекурентні нейронні мережі;
- згорткові та глибинні нейронні мережі;
- радіально-базисні;
- самоорганізовані карти Кохонена.

Далі будуть представлені рекурентні нейронні мережі, оскільки "Всякий раз, коли існує послідовність даних, і тимчасова динаміка, яка пов'язує дані, важливіша за просторовий зміст кожного окремого кадру", – Лекс Фрідман (MIT).

Рекурентна нейронна мережа (RNN) - це клас штучних нейронних мереж, де зв'язки між вузлами утворюють спрямований графік уздовж часової послідовності. Це дозволяє йому демонструвати часову динамічну поведінку. Похідні від нейронних мереж прямого зв'язку, RNN можуть використовувати свій внутрішній стан (пам'ять) для обробки послідовностей змінної довжини входів.

RNN є особливим видом нейронних мереж, які призначені для ефективної роботи із послідовними даними. Такі типи даних включають в себе часові ряди (списки значень деяких параметрів за певний проміжок часу), текстові документи, які можна розглядати як послідовність слів або ауді, які можна розглядати як послідовність звукових частот в часі.

RNN є рекурентні по своїй природі, оскільки виконують ту ж саму функцію для кожного вводу даних, в той час, як вивід нинішнього вводу

залежить від останнього обчислення. Для прийняття рішення він враховує нинішній ввід і вивід, який йому відомий із попереднього вводу. Комірки, які являються функцією входних даних із попередніх часових кроків також відомі як комірки пам'яті. На відміну, від нейронних мереж із прямим зв'язком, RNN можуть використовувати свій внутрішній стан (пам'ять) для обробки послідовних входних даних. Інші нейронні мережі мають незалежні одне від одного входи, на відміну, від рекурентних, у яких всі входи взаємопов'язані (Рисунок 2.1).

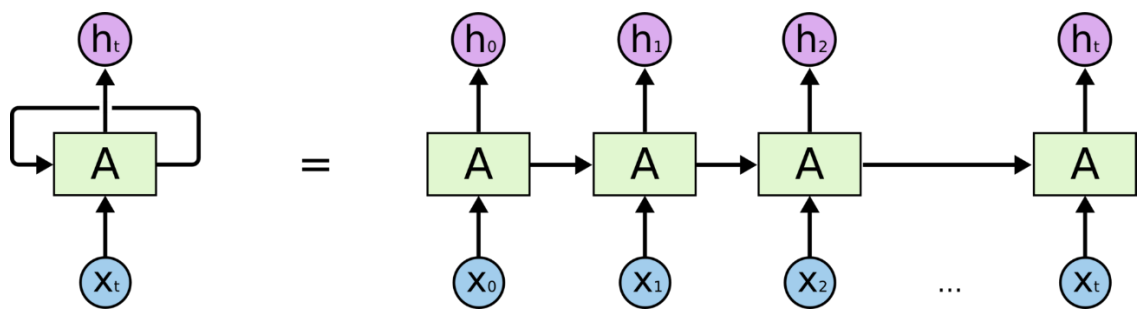


Рисунок 2.1 – Схема розгорнутої RNN

Спочатку він бере x_0 з послідовності введення, а потім виводить h_0 , який разом з x_1 є входом для наступного кроку. Отже, h_0 і x_1 - це входні дані для наступного кроку. Аналогічно, h_1 з наступного - це вхід з x_2 для наступного кроку і так далі. Таким чином, він постійно запам'ятовує контекст під час тренування [7].

Формула для поточного стану:

$$h_t = f(h_{t-1}, x_t)$$

До недоліків рекурентних нейронних мереж можна віднести:

1. Проблеми із затуханням або вибуху градієнту.
2. Навчання RNN – складний процес.

3. RNN не може оброблювати дуже довгі посліовності, якщо в якості функції активації використовується \tanh або relu .

Частину цих недоліків усуває нейронна мережа з довгою короткостроковою пам'яттю (LSTM).

2.2.2 Нейронна мережа з довгою короткостроковою пам'яттю (LSTM)

Мережі з довгої короткостроковою пам'яттю (LSTM) представляють собою модифіковану версію рекурентних нейронних мереж, яка спрощує запам'ятовування минулих даних в пам'яті. Тут вирішена проблема затухання градієнта RNN. LSTM добре підходить для класифікації, обробки і прогнозування часових рядів з урахуванням часових лагів невідомої тривалості. Він навчає модель за допомогою зворотного поширення.

LSTM мають більш складну клітинну структуру, ніж звичайний рекурентний нейрон, що дозволяє їм краще регулювати, як ефективно вчитися чи забувати з різних джерел вхідних даних.

Ключем до LSTM є стан комірки (пам'ять комірки), горизонтальна лінія, що проходить через верхню частину діаграми, через яку проходить інформація, і внутрішній механізм, що називається воротами, які можуть регулювати потік інформації. Стан клітини нагадує конвеєр. Він проходить прямо по всьому ланцюжку, лише з деякими незначними лінійними взаємодіями. Стан комірки в основному кодує інформацію про входи (відповідну інформацію), які спостерігались до цього кроку (на кожному кроці).

У мережі LSTM присутні три шлюзи (Рисунок 2.2).

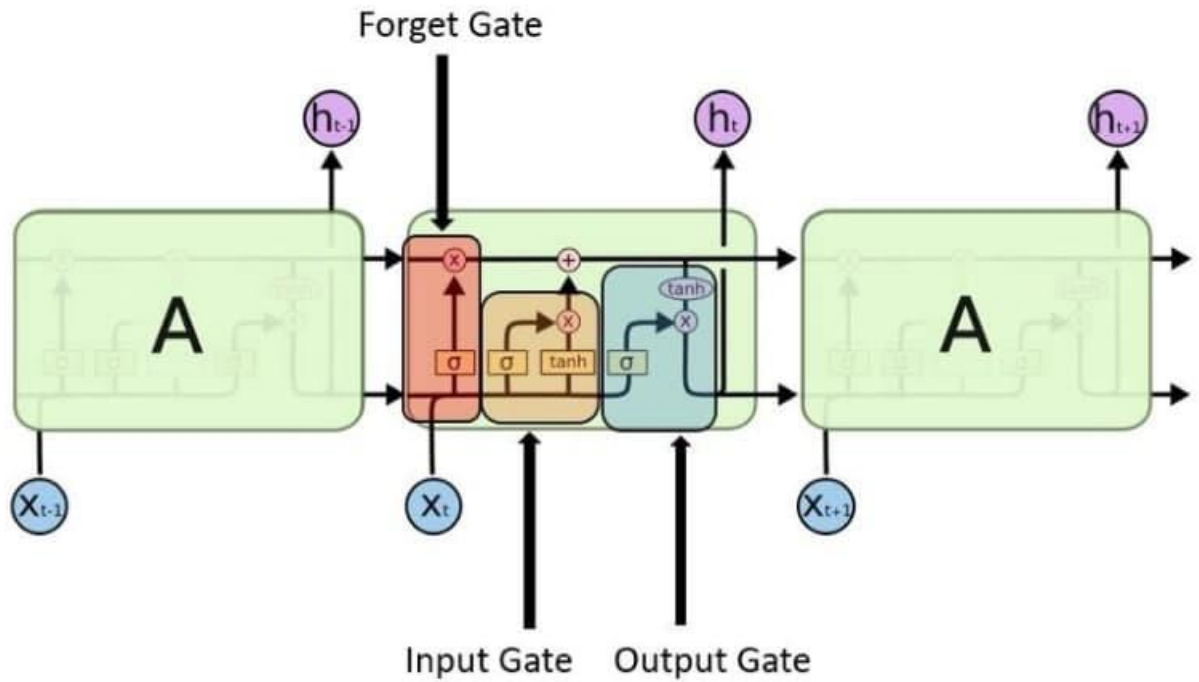


Рисунок 2.2 – Схема LSTM

1. Forget gate — за допомогою цього етапу мережа вирішує, яку інформацію необхідно забути в поточному елементі (cell) за допомогою forget gate layer. Це визначається сигмоїдною функцією, яка дивиться на попередній стан h_{t-1} і інформацію з поточного часового кроку x_t та виводить число від 0 (пропустити) до 1 (залишити це) для кожного числа в стані осередку C_{t-1} .

$$f_t = \sigma (W_t \cdot [h_{t-1}, x_t] + b_f),$$

де W_t — вагові коефіцієнти.

2. Input gate — на даному етапі приймається рішення чи хочемо зберігати інформацію в поточному стані чи ні, тобто, яке значення з входу слід використовувати для зміни пам'яті. Для цього мережа використовує стан з попереднього кроку та

значення послідовності в поточному часовому кроці та подаємо на сигмоїду. Складається з двох шарів : шар сигмоїдної функції, який вирішує які значення можуть бути оновлені та шар \tanh , який створює вектор нових значень, які є кандидатом на оновлення елементів.

$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh (W_c \cdot [h_{t-1}, x_t] + b_c),$$

де \tilde{C}_t – вектор нових значень кандидатів.

3. Output gate — на даному етапі мережа застосовує для оновлення стану те, що було розраховано та вирішено на минулих кроках, для створення нового стану. Де власне приймаються рішення, що буде повернуто поточним елементом. Для цього спочатку використовується шар сигмоїдної функції, на якому вирішується яку частину поточного стану буде повернуто. Потім розраховується значення кандидатів, що можна повернути. І наприкінці поєднуються результати всіх шарів і повертається тільки те, що можна повернути.

$$o_t = \sigma (W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh (C_t)$$

Ідея полягає в тому, як процес навчання базується на контексті (пам'яті). Ви забуваєте, дізнаєтесь і витягуєте частину цього для наступного туру, але в наступному турі ви знову повторюєте той самий процес. В основному, ми намагаємось імітувати, як людський мозок намагається пізнати речі за допомогою механізму внутрішніх воріт LSTM (і це може бути

не обов'язково правдою, ми просто намагаємось застосовувати різні підходи).

Висновки до розділу 2

У даному розділі було розглянуто два підходи до моделювання часових рядів: регресійний підхід та за допомогою рекурентних нейронних мереж, зокрема LSTM. Залежно від особливостей даних кожний із цих підходів має свої переваги та недоліки.

Загальним недоліком статистичних моделей є складність вибору типу моделі і підбору її параметрів. Все це істотно збільшує суб'єктивний внесок учасників процесу аналізу і прогнозування часових рядів. Таким чином, результат аналізу і прогнозування часових рядів залежить як від кваліфікації аналітика в предметній галузі, так і від його кваліфікації в методах аналізу. Апарат нейронних мереж мінімізує участь аналітика в формуванні моделі часового ряду, так як здатність нейромережевих моделей до навчання дозволяє виявити приховані взаємозв'язки і закономірності між даними, а алгоритми навчання адаптують вагові коефіцієнти відповідно до структури даних, представлених для навчання.

З іншої сторони, навчання моделей за допомогою нейронних мереж мають і свої недоліки в порівнянні перед авторегресійними моделями. Наприклад, мережі можуть мати проблеми із затуханням або ж вибухом градієнту, встановлення необхідного значення коефіцієнту швидкості навчання емпіричним шляхом, складність та багатошаровість самої структури НМ, а також потреба у великій обчислювальній потужності. У той час, коли структура авторегресійних моделей і процес їх створення є порівняно легшим і вимагає менше обчислювальних можливостей, інколи статистичний ряд, що

має мало даних доцільніше прогнозувати за допомогою таких простих моделей як AR та ARMA.

Чим більша послідовність – тим глибша рекурентна нейронна мережа, тобто виникають проблеми з довгостроковими послідовностями. Дійсно на практиці нейронна мережа не може навчитись працювати з досить довгостроковими зв'язками. Причиною цього є градієнтне затухання, де з кожним шаром сила(інтенсивність) градієнту зменшується, тобто рекурентна НМ буде втрачати більш дальні контекст.

Отже, метод для прогнозування часових рядів необхідно обирати враховуючи особливості статистичних даних.

РОЗДІЛ 3

МЕТОДИКА МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ І ЇЇ СТРУКТУРА

3.1 Формування структури моделі

Перш ніж розпочати роботу з математичними методами прогнозування та побудовою їх моделей необхідно якісно сформувати структуру моделі. Адже під час алгебраїчних та статистичних перетворень дана модель може поводити себе некоректно та давати результати, які не відповідають дійсності. На даному етапі необхідно скористатися всіма наданими джерелами інформації про процеси із метою вивчення їх особливостей.

Оскільки модель може містити збурення, що впливають на процес, визначення наявності затримок на якісному та кількісному рівні, якщо це можливо визначення порядку процесу, а також наявності нелінійності і її характер. Якщо дослідження відбувається економічних процесів, то необхідно виявити чи наявна сезонність та чи присутній тренд, тобто зробити аналіз на інтегрованість. Виявивши в експерименті суттєву зміну рівня коливань на інтервалах часових рядів, зробити висновок про гетероскедастичність, а також перевірити гіпотезу про можливість введення коінтегрованості змінних. У результаті проведеного аналізу процесу можна побудувати ймовірну структуру моделі, яка і буде використовуватись в наступних засобах прогнозування.

На основі даних міркувань можна сформувати етапи реалізації системи прогнозування [8]:

Етап 1. Вибір адекватної моделі

На першому етапі необхідно обрати модель, яка з найбільшим рівнем адекватності буде описувати динамічний ряд.

Етап 2. Вибір методів прогнозування

На другому потрібно обрати для часового ряду, що досліджується, кращий метод прогнозування.

Етап 3. Попередня обробка даних

На даному етапі відбувається вибір методів для виконання обробки досліджуваних даних, яка може складатися з: нормування, логарифмування, заповнення пропусків або ж навпаки згладжування.

Етап 4. Оцінювання параметрів моделі

Для лінійних моделей:

- метод найменших квадратів (МНК) та рекурсивний МНК;
- метод максимальної правдоподібності (ММП) та рекурсивний ММП;
- метод допоміжної змінної (МДЗ).

Для нелінійних моделей:

- нелінійний метод найменших квадратів (НМНК);
- метод максимальної правдоподібності (ММП) та рекурсивний ММП;
- узагальнений метод моментів (УММ);
- метод Монте Карло для Маяковських ланцюгів.

Етап 5. Вибір моделей із існуючих кандидатів

На цьому етапі відбираються моделі, які краще проявили себе на обраному часовому ряді та виконується формування нових критеріїв добору за допомогою розглянутих критеріїв адекватності.

Етап 6. Оцінка якості прогнозу

Даний етап являє собою вибір методів обчислення якості прогнозування моделей.

3.1.1 Аналіз нелінійності

Варто зауважити, що перевірка на нелінійність може виконуватись на основі різних критеріїв, проте необхідно враховувати про всі їхні особливості. Результати різноманітних варіацій функцій, що мають характер нелінійності, не завжди є вдалимими.

У економічних процесах допускається використання дисперсійного методу, що допомагає визначити наявність нелінійності, за допомогою функції:

$$\Psi_{zu}(t_1, t_2) = E_{u(t_2)}[E_{z(t_1)}[z(t_1)|u(t_2)] - E_{z(t_1)}[z(t_1)]]^2 ,$$

розв'язання даної функції відбувається за допомогою складного інтегрального рівняння.

Вирішити проблему нелінійності прогнозування процесів можна також за допомогою статистичного загального тесту Фішера (f-тест) [16]:

$$\hat{F} = \frac{\frac{1}{k-2} \sum_{i=1}^k \sum_{j=1}^{n_j} n_i (\bar{y}_i - \hat{y}_{ij})^2}{\frac{1}{k-2} \sum_{i=1}^k \sum_{j=1}^{n_j} (y_{ij} - \bar{y}_i)^2} \quad (3.1)$$

де k – число груп даних;
 n_i – кількість вимірів у групі;
 \bar{y}_i – групове середнє;
 \hat{y}_{ij} – оцінка по прямій регресії;
 n – загальне число вимірів.

Перефразувавши даний статистичний набір (3.1), отримаємо наступне відношення:

$$\hat{F} = \frac{\text{Відхилення середніх значень від прямої регресії}}{\text{Відхилення значень } y(k) \text{ від групових середніх}}$$

Припущення про лінійність процесу вважається хибних, якщо статистичний набір \hat{F} із ступенями свободи рівними $\nu_1 = k - 2$, $\nu_2 = n - k$ дорівнює або більший рівня значущості[8].

3.1.2 Перевірка на стаціонарність

Проблема стаціонарності процесу вирішується за допомогою розширеного тесту Дікі-Фуллера. Особливістю даного тесту є те, що значення залежної змінної з великими значеннями лагу вводиться в рівняння регресії, якого вистачить, щоб в тестові не застосовувати автокореляційні залишки. Дане рівняння набуває вигляду:

$$\Delta y(k) = a_0 + by(k-1) + c_1\Delta y(k-1) + c_2\Delta y(k-2) + \dots + c_n\Delta y(k-n) + \varepsilon(k).$$

Середнє значення і компонент, за допомогою якого описується тренд утворюють окремий вид моделі, що тестується, та утворює форму критерію значимості.

Рівняння, яке тестується:

$$\Delta y(k) = by(k-1) + c_1\Delta y(k-1) + c_2\Delta y(k-2) + \dots + c_n\Delta y(k-n) + \varepsilon(k),$$

де середнє значення рівне нулю та гіпотеза формулюється наступним чином:

$H_0: b = 0$ – ряд нестационарний;

$H_1: b < 0$ – ряд стаціонарний.

Якщо статистика $\frac{b}{SE_b}$ має значення менше від нуля та менше за критичне значення, яке було отримано із таблиці Діккі-Фуллера, то нульова гіпотеза не розглядається. Критичні значення рівнів значимості відповідно дорівнюють $\alpha = 1$ – -2,85 і $\alpha = 5$ – -1,95.

Якщо нульову гіпотезу є задовільною, то ряд $\{y(k)\}$ – являє собою випадкове блукання без зсуву, в рівнянні це константи.

У загальному вигляді цей критерій можна подати у вигляді обчислення модифікованого критичного значення з урахуванням розміру вибірки, що рівний N , та обчислюється за формулою:

$$\tau_{\infty} + \frac{\tau_1}{N} + \frac{\tau_2}{N^2}$$

де $\tau_{\infty} = -2.57$ ($\alpha = 1$) чи $\tau_{\infty} = -1.94$ ($\alpha = 5$);

$\tau_1 = -1.96$ ($\alpha = 1$) чи $\tau_1 = -0.398$ ($\alpha = 5$);

$\tau_2 = -10.04$ ($\alpha = 1$) чи $\tau_2 = 0$ ($\alpha = 5$) (значення τ табульовані Маккінномом, 1991).

Перевірка рівняння $\Delta y(k) = a_0 + by(k-1) + \varepsilon(k)$ з врахуванням можливої автокореляції залишків (як це було показано вище) базується на використанні того ж статистичного критерію, що і для рівняння без середнього, і тієї ж формули критичних значень, але з урахуванням наступних значень τ :

$\tau_{\infty} = -3.43$ ($\alpha = 1$) чи $\tau_{\infty} = -2.86$ ($\alpha = 5$);

$\tau_1 = -6.00$ ($\alpha = 1$) чи $\tau_1 = -2.74$ ($\alpha = 5$);

$\tau_2 = -29.25$ ($\alpha = 1$) чи $\tau_2 = -8.36$ ($\alpha = 5$).

При наявності середнього та тренду застосовується така ж процедура, що і вище, але при наступних значеннях τ :

$$\begin{aligned}\tau_{\infty} &= -3.96 \ (\alpha = 1) \text{ чи } \tau_{\infty} = -3.41 \ (\alpha = 5); \\ \tau_1 &= -8.35 \ (\alpha = 1) \text{ чи } \tau_1 = -4.04 \ (\alpha = 5); \\ \tau_2 &= -47.44 \ (\alpha = 1) \text{ чи } \tau_2 = -17.83 \ (\alpha = 5).\end{aligned}$$

Розглянемо розширений тест Дікі-Фуллера.

Необхідно побудувати рівняння регресії, для використання тесту Дікі-Фуллерах[9]:

$$\Delta y(k) = a_0 + a_1 k + b y(k-1) + \sum_{i=1}^p c_i \Delta y(k-i) + \varepsilon(k), \quad (3.2)$$

де a_0, a_1, b, c_i – невідомі коефіцієнти регресії.

Вище вказане рівняння (3.1) можна застосувати для реалізації тесту Діккі-Фуллера за умови, що коефіцієнти $c_i = 0, i = 1, 2, \dots, p$, в іншому випадку потрібно використовувати загальний тест Дікка-Фуллера. У сфері економіки краще використовувати загальний тест Дікка-Фуллера із кількістю значень p , які мали затримку в часі, меншою, ніж десять відсотків від числа спостережень, де $p < 0.1N$, де N – довжина(потужність) часового ряду. Варто зауважити, що важливим моментом, коли використовується загальний та звичайний тест Діккі-Фуллера, є правильне формування структури моделі, адже важливу роль відіграє наявність параметрів a_0 і $a_1 k$.

Можна сформулювати евристичне правило, яке допоможе прискорити процес розв'язку даної задачі, а саме з візуалізації графіка зробити аналіз про наявність тренду. Якщо ж тренд відсутній, то в отриману модель (3.2) потрібно включити тільки вільний член a_0 , що є перетином. В іншому

випадку, коли аналіз графіку вказує на присутність тренду, то в модель (3.2) вводяться наступні параметри a_0 і $a_1 k$.

Можна сформулювати наступні гіпотези на основі моделі (3.2):

$H_0: b = 0$ – ряд нестационарний: $y(k) \sim I(int)$, $int > 0$;

$H_1: b < 0$ – ряд стаціонарний: $\{y(k)\} \sim I(0)$, $int = 0$.

H_0 не розглядається, якщо маємо наступну нерівність оцінки коефіцієнта $\hat{b} < 0$ та виконаємо тестування на наявність одиничного кореня за допомогою обчислення τ – статистика Маккіннона, яка за абсолютною величиною є більшою за величину критичного значення даної статистики при регульованому рівні значущості α . [16]

Це можна записати наступним чином:

$$|\tau| = \left| \frac{\hat{b}}{SE_{\hat{b}}} \right| \geq |\tau_{крит}|$$

з рівнем значущості α , де $SE_{\hat{b}}$ є стандартною похибкою оцінки \hat{b} .

3.1.3 Наявність коінтегрованості

Інтегрованість ряду – це перевірка динаміку ряду, який досліджується, на наявність тренду. Тестування можна здійснити певною кількістю способів, які будуть наведені в порядку збільшення рівня складності реалізації:

- а) Графічний метод являє собою візуальне виявлення тренду на основі продемонстрованого графіку. Побудова графіку виконується наступним чином: по осі абсцис відкладається час, а по осі ординат

значення ряду. За характером поведінки кривої можна зробити про присутність або відсутність тренду часового ряду.

- б) Метод середніх використовується після поділу досліджуваного ряду на два рівних підряди, для кожного з яких визначається середня величина \bar{Y}_1 та \bar{Y}_2 . У результаті якщо знайдені значення відрізняються більше ніж на 10%, то можна констатувати факт наявності тренду в динамічному ряді.
- в) Метод Стюарта та Кокса в дечому схожий із попередньо розглянутим, проте значною відмінністю є поділ того ж таки ряду на 3 рівні за кількістю рівнів групи. Особливої уваги заслуговує порівняння рівнів 1-ї та 3-ї груп. Якщо часовий ряд не можна порівну розділити на 3 частини, тобто кількість рівнів не ділиться на 3, то необхідно вилучити або ж додати рівень, якого не вистачає.
- г) У методі Мура та Валліса наявність тренду підтверджується тільки в тому випадку, якщо динамічний ряд не містить взагалі або ж містить в допустимій кількості фази, тобто зміну знаку під час ідентифікації абсолютного показника перетворення способом прив'язки.
- д) Метод наборів використовується, коли кожний рівень ряду належить одному з двох типів. Наприклад, першому типові належить частина, яка менше середнього значення або ж медіани, а другому типові більше цих значень. Потім встановлюється, в уже створених типах послідовностях, кількість наборів R . Вони є послідовностями одного ж типу ряду, що межують з рівнями послідовності іншого типу. Якщо ряд не має загальну тенденцію росту або зниження рівнів, то кількість наборів є випадковою величиною з нормальним законом розподілу, за умови, що $n > 30$, або ж, якщо $n < 30$, то використовуємо розподіл Стюдента. У результаті, якщо в змінах рівнів немає закономірностей, то випадкова величина належить довірчому інтервалу:

$$(\bar{R} - t\sigma) \leq R \leq (\bar{R} + t\sigma) \quad (3.3)$$

де t – довірчий коефіцієнт для допустимого рівня ймовірності в нормальному розподілі, а в розподілі Стюдента із ступенем свободи рівним $k = (n - 1)$.

\bar{R} – середня кількість наборів в ряді, що знаходиться за формулою:

$$\bar{R} = \frac{n+1}{2} \quad (3.4)$$

σ – середнє квадратичне відхилення кількості наборів в ряді, що розраховується за формулою:

$$\sigma = \frac{\sqrt{n-1}}{2}$$

Підставимо виведення показників (3.4) та (3.5) в довірчий інтервал (3.3) та отримаємо наступне перетворення:

$$\frac{(n + 1 - t\sqrt{n - 1})}{2} \leq R \leq \frac{(n + 1 + t\sqrt{n - 1})}{2}$$

Отже, якщо встановлена кількість наборів ряду не входить в довірчий інтервал, то у динамічному ряді наявний тренд, якщо значення R належить довірчому інтервалу, то можна констатувати відсутність тренду[16].

3.1.4 Аналіз на гетероскедастичність

Існує декілька тестів за допомогою яких можна дослідити часовий ряд на наявність гетероскедастичності.

Розглянемо Тест Бройша-Пагана (Годфрі).

Маємо наступну лінійну регресію:

$$y(k) = X^T(k)\beta + \varepsilon(k), \quad (3.6)$$

де $X^T(k) = [1 \ x_2(k) \ x_3(k) \ \dots \ x_r(k)]$.

Нехай гетероскедастичність приймає наступний вигляд:

$$\begin{aligned} E[\varepsilon(k)] &= 0, \quad \forall k, \\ \text{var}[\varepsilon(k)] &= E[\varepsilon^2(k)] = \sigma_\varepsilon^2 = h(\alpha z^T(k)), \end{aligned} \quad (3.7)$$

де $z^T(k) = [1 \ z_2(k) \ z_3(k) \ \dots \ z_p(k)]$ – вектор відомих змінних;

$\alpha = [\alpha_1 \ \alpha_2 \ \alpha_3 \ \dots \ \alpha_p]$ – вектор невідомих коефіцієнтів;

$h(\cdot)$ – довільна невизначена функція, що приймає лише значення більші нуля.

Нульова гіпотеза гомоскедастичності:

$H_0 : \alpha_2 = \alpha_3 = \dots = \alpha_p = 0$, що означає $\sigma_\varepsilon^2 = h(\alpha_1) = \text{const.}$

За цією нульовою гіпотезою можна оцінювати коефіцієнти виразу (3.6) з використанням методу найменших квадратів, припустивши, що розподіл збурень у правій частині рівняння є нормальним. Процедуру застосування тесту на гетероскедастичність можна легко зобразити у вигляді нескладного алгоритму[16]:

- 1) Спочатку необхідно здійснити оцінку параметрів початкової моделі (3.6), використовуючи простий МНК, а також сформувати масив

залишків $e(k) = y(k) - X^T(k)\beta$ та обчислити дисперсії $\sigma_e^2 = \sigma_e^2 = N^{-1} \sum e^2(k)$;

- 2) Знайти оцінки регресії $e^2(k)/\sigma_e^2$ на $z(k)$ за допомогою ЗМНК та обчислити значення похибки ESS за формулою $ESS = \beta^T X^T X \beta - N\mu_y^2$, де μ_y^2 – середнє значення послідовності $\{y(k)\}$;
- 3) Застосовуючи нуль-гіпотезу $H_0: \frac{1}{2}ESS \leftrightarrow \chi^2(p-1)$. Отже, гіпотеза стосовно гомоскедастичності не береться до уваги, якщо $ESS/2$ перевищує вибране критичне значення із розподілу χ^2 ;
- 4) Асимптотично еквівалентним підходом, який є легше реалізувати, є оцінювання регресії $e^2(k)$ на $z(k)$. Величина NR^2 , обчислена для цієї регресії, буде мати в асимптотичний розділ $\chi^2(p-1)$ при розгляді нульової гіпотези.

Варто зауважити, що зв'язок між різними видами похибок регресійної моделі визначається за виразом: $(y^T y - N\mu_y^2) = (\beta^T X^T X \beta - N\mu_y^2) + e^T e$, або $TSS = ESS + RSS$,

де TSS – загальна похибка регресії;

RSS – сума квадратів похибок моделі.

Використовуючи продемонстрований тест необхідно знати змінні z , які створюють гетероскедастичність, але немає потреба завчасно знати функціональну структуру гетероскедастичності. У дослідженнях кандидати в змінні z можуть бути вибрані з векторів регресорів $x(k)$. Якщо дане припущення підтверджується, то алгоритм застосування даного теста відповідає алгоритму тесту Уайта[8].

Розглянемо Тест Уайта.

Для тестування гетероскедастичності у цьому тесті потрібно сформулювати додаткову модель регресії для квадратів залишків, що будуть генеруватись з використанням методу найменших квадратів. Дана модель регресія буде складатися з ненадлишкових регресорів на всій області регресорів, що містять квадрати, взаємні добутки регресорів та власне їх, а

також модель міститиме константу в правій частині. Побудуємо регресію у вигляді:

$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + \varepsilon(k),$$

де $[1 \ x_1 \ x_2]^T$ - вектором незалежних змінних.

Множина всіх регресорів, їх квадратів та взаємних добутків, ненадлишкових змінних матиме наступну форму: $[1 \ x_1 \ x_2 \ x_1^2 \ x_2^2 \ x_1 x_2]$.

Якщо модель має гетероскидастичні процеси, то $NR^2 \leftrightarrow \chi^2(q)$. Тобто NR^2 може мати асимптотичний розподіл з параметром q , який означає кількість регресорів без константи і в даному випадку буде дорівнювати 5, $\chi^2(5)$. Даний тест може виявити гетероскедастичність часового ряду, проте не демонструє структуру початкової моделі й спосіб знаходження її параметрів. Але оцінку параметрів можна знайти за допомогою методу найменших квадратів.

Також значним недоліком даного тесту є те, що розподіл χ^2 може мати велику кількість рівнів свободи, що сприятиме значному збільшенню параметру q і відповідно зменшуватиме якість даного тесту. Іноді за рахунок вилучення взаємних добутків із загальної множини регресорів та шляхом вводу в регресію їх квадратів, зменшують значення параметра q [8].

Також існує тест Голдфелда-Квандта, який використовують у випадках, коли є одна змінна, зокрема з числа регресорів, що породжує гетероскедастичність. Якість даного тесту залежить від об'єму спостережень, що не розглядаються в дослідженні.

3.2 Критерії вибору кращої моделі та прогнозу

Перш ніж формувати структуру математичної моделі прогнозування необхідно вибрати, яка з моделей буде краще описувати характер поведінки динамічних процесів, адже процеси в різних галузях їхнього прояву мають різні особливості. Для вирішення даної проблеми існують різноманітні критерії адекватності моделі, які допомагають вірно обрати модель прогнозування та оцінити її актуальність щодо досліджуваного процесу. До відомих критеріїв належать: статистичні параметри (t – статистика Стюдента, коефіцієнт детермінації R^2 , сума квадратів похибок моделі(SSE)), інформаційний критерій Акайке (AIC), статистика Дарбіна-Уотсона(DW), статистика Фішера, коефіцієнт Тейла (Theile).

Після того, як була вибрана вже модель і побудований прогноз необхідно визначити чи є цей прогноз доцільний. Отримавши прогноз різними математичними моделями, кожна модель має свої переваги та недоліки в інформативності зробленого прогнозування. Тому існують оцінки точності прогнозу, що і характеризують його: середньоквадратична похибка(СКП), середня похибка прогнозу(СП), середня похибка в процентах(СПП), середня абсолютна похибка у процентах(СаПП), максимальна абсолютна похибка(МАП), мінімальна абсолютна похибка (MiАП).

3.2.1 Критерії адекватності моделі

Критерії адекватності моделі дозволяють оцінити окремо значущість коефіцієнтів математичної моделі в статистичному сенсі, визначити

інтегральну похибку моделі стосовно вихідного часового ряду, встановити наявність кореляції між значеннями похибки моделі, адже вони мають бути не корельованими, а також визначити ступінь адекватності моделі фізичному процесу в цілому. Розглянемо деякі з них.

Розглянемо коефіцієнт детермінації R^2 .

Оскільки мірою інформативності часового ряду найчастіше використовують дисперсію, то R^2 базується саме на цій ідеї. Обчислюється коефіцієнт детермінації за формулою:

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} = 1 - \frac{SSE}{SST}$$

де $\text{var}(\hat{y})$ – дисперсія частини часового ряду основної змінної рівняння;

$\text{var}(y)$ – вибіркова дисперсія цієї змінної;

$SSE = \sum_{k=1}^N [y(k) - \hat{y}(k)]^2$ – сума квадратів похибок (залишків) моделі;

$SST = \sum_{k=1}^N [y(k) - \bar{y}]^2$ – загальна сума квадратів \bar{y} – середнє

значення.

Модель вважається кращою, якщо $R^2 \rightarrow 1$.

Розглянемо критерій, що являється сумою квадратів похибок моделі (SSE)

Обчислюється за формулою:

$$SSE = \sum_{k=1}^N e^2(k) = \sum_{k=1}^N [\hat{y}(k) - y(k)]^2 \rightarrow \min$$

де $\hat{y}(k) = \hat{a}_0 + \hat{a}_1 \hat{y}(k-1) + \hat{a}_2 \hat{y}(k-2) + \hat{b}_1 x(k) + b_2 z(k)$;

$y(k)$ – вимірювання;

N – довжина вибірки.

Отже, для отримання найкращої моделі з усіх можливих кандидатів потрібно обрати ту, в якій значення $\sum e^2(k)$ є найменшим.

Розглянемо критерій Акайке (AIC)

Даний критерій бере до уваги $\sum e^2(k)$, N-вимірів та число параметрів моделі. Обчислюється за наступною формулою:

$$AIC = N \ln \left(\sum_{k=1}^N e^2(k) \right) + 2n$$

де $n = p + q + 1$ – число параметрів моделі, які оцінюються за допомогою статистичних даних (p - число параметрів авто регресійної частини моделі; q - число параметрів ковзного середнього; одиниця з'являється тоді, коли оцінюється зміщення (або перетин), тобто a_0).

Оскільки даний метод враховує значення $\sum e^2(k)$, то при виборі кращої моделі-кандидата, необхідно прагнути мінімального значення AIC . Інформативнішим його робить наявність у формулі таких значень як N-довжина вибірки та число оцінюваних параметрів[16].

Розглянемо такий критерій, як Статистика Дарбіна-Уотсона (DW).

Обчислюється за формулою:

$$DW = 2 - 2\rho,$$

де $\rho = \frac{E[e(k)e(k-1)]}{\sigma_e^2}$ – коефіцієнт кореляції між сусідніми значеннями похибки;

σ_e^2 – дисперсія послідовності похибок $\{e(k)\}$.

Таким чином, при повній відсутності кореляції між похибками $DW = 2$ – це ідеальне значення. Граничними значеннями для DW є 0 (при $\rho = 1$) та +4 (при $\rho = -1$).

Розглянемо коефіцієнт Тейла (Theile)

Обчислюється за формулою:

$$U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i)^2} + \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i)^2}}$$

Маємо допустимі значення коефіцієнта $0 \leq U \leq 1$. Модель не може використовуватись у прогнозуванні при $U = 1$, інакше, коли $U = 0$ це означає, що ряд прогнозу співпадає з реальними рядом, тобто в такому випадку модель найкраще описує реальний процес.

3.2.2 Оцінювання точності прогнозу

Після побудови та виборі найкращої моделі, необхідно обчислити оцінку прогнозу ряду та виявити поведінку ряду.

1) Середньоквадратична похибка

$$\text{СКП} = \sqrt{\frac{1}{S} \sum_{s=1}^S (y(k+s) - \hat{y}(k+s, k))^2}$$

2) Середня похибка прогнозу:

$$\text{СП} = \frac{1}{S} \sum_{s=1}^S y(k+s) - \hat{y}(k+s, k)$$

3) Середня похибка в процентах:

$$\text{СПП} = \frac{1}{S} \sum_{i=1}^S \frac{y(k+s) - \hat{y}(k+s, k)}{y(k+s)} \times 100\%$$

4) Середня абсолютна похибка у процентах:

$$\text{АСПП} = \frac{1}{S} \sum_{i=1}^S \frac{|y(k+s) - \hat{y}(k+s, k)|}{|y(k+s)|} \times 100\%$$

5) Максимальна абсолютна похибка:

$$\text{МАП} = \max\{|y(k+1) - \hat{y}(k+1, k)|, \dots, |y(k+s) - \hat{y}(k+s, k)|\}.$$

6) Мінімальна абсолютна похибка:

$$\text{МіАП} = \min\{|y(k+1) - \hat{y}(k+1, k)|, \dots, |y(k+s) - \hat{y}(k+s, k)|\}$$

Висновки до розділу 3

Даний розділ містить поетапну структуру процесу моделювання та прогнозування часових рядів, а саме шість етапів:

- вибір адекватної моделі;
- вибір методу прогнозування;
- попередня обробка даних;
- оцінка параметрів моделі;
- вибір найкращих кандидатів;

— оцінка якості прогнозу.

Продемонстровано можливі особливості часових рядів та представлено тести для їх перевірки на лінійність, стаціонарність, гетероскедастичність та інтегрованість, які особливо характерні для даних, що мають збурення та сезонний характер, суттєву зміну коливань на часових інтервалах, притаманні економічним та екологічним процесам. Також наведені найбільш використовувані критерії адекватності моделі та оцінки якості прогнозу.

РОЗДІЛ 4

ПОБУДОВА МАТЕМАТИЧНИХ МОДЕЛЕЙ ЕКОНОМІЧНИХ ТА ЕКОЛОГІЧНИХ ПРОЦЕСІВ

4.1 Програмна реалізація та її архітектура

Результатом мого магістерського дослідження є реалізований програмний продукт для створення математичних моделей та прогнозування динамічних часових рядів, апробований на статистичних даних деяких економічних та екологічних процесів.

Метою створення програмного продукту є система за належним функціоналом, за допомогою якої користувач без спеціальної підготовки міг би виконувати дослідження особливостей динаміки часових рядів для яких характерні збурення, сезонність, нелінійність та коінтегрованість. Представлена робота була побудована на основі статистичних даних природного стану довкілля та економічної ситуації для аналізу проблем, що нині існують у світі та засобів для їх запобігання в майбутньому.

Далі представлена функціональна схема системи (Рисунок 4.1):

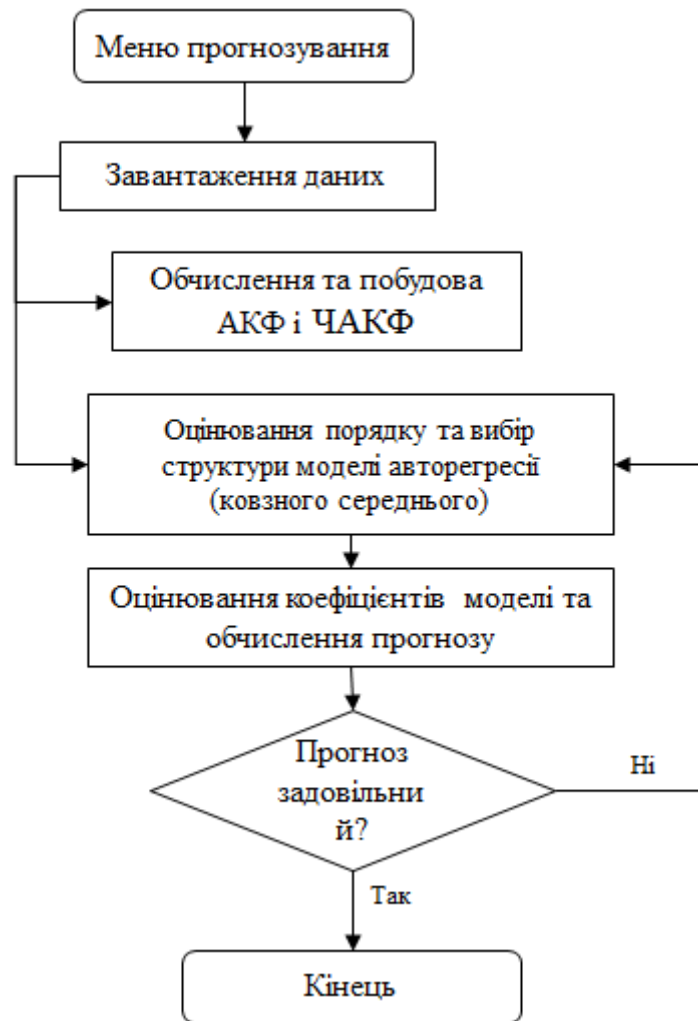


Рисунок 4.1 – Функціональна схема системи

4.2 Аналіз вибору інструментальної платформи для реалізації системи

Під час написання роботи для вибору середовища створення програмного продукту було розглянуто наступні варіанти: Python, R, Matlab та C#. Виконуючи огляд функціональних можливостей та можливу візуалізацію результатів дослідження було прийнято рішення віддати перевагу Python, робота була реалізована в програмному середовищі Jupyter Notebook з версією Python 3.7.4.

Python є однією з найбільш популярних мов програмування найвищого рівня, що має строгу динамічну типізацію. Вона має ряд переваг, навіть тільки можливість її використання в режимі діалогу, може приваблювати будь-якого користувача, оскільки це дуже зручно при роботі з дослідженнями та виконанням експериментів. Python є зручним для розв'язування математичних задач та виконання моделювання, а також дане програмне середовище має інформативну візуалізацію різноманітного типу, приваблює ще й те, що немає необхідності попереднього виділення пам'яті та завчасного задання розмірності і типу даних при роботі з матрицями та великими масивами даних.

Однією з головних переваг серед інших середовищ є наявність зручної бібліотеки Statsmodels, вона вже має стандартні підходи математичного моделювання та виконання прогнозу. Дана бібліотека є пакетним продуктом Python, що дає можливість дуже легко аналізувати та проводити дослідження над даними, виконувати статистичні тести та проводити оцінювання побудованих моделей.

Python містить уже створений модуль статистики SciPy, який має великий перелік вже підготовленої необхідної для експериментів статистики, побудови графічних функцій та візуалізації статистичних результатів, що не може не приваблювати користувача. Statsmodels є частиною наукового стеку Python, який орієнтований на аналіз даних, наукових даних і статистику. Statsmodels побудований поверх чисельних бібліотек NumPy і SciPy, також вона інтегрується з Pandas для обробки даних і використовує Patsy для інтерфейсу R-подібної формули. Бібліотека Matplotlib надає можливість користувачеві будувати графічні функції. Statsmodels надає статистичний сервер для інших бібліотек Python. Statsmodels є вільним програмним забезпеченням, що випускається під ліцензією модифікованої BSD (3-клаузи).

4.3 Побудова математичних моделей та прогнозування на основі статистичних даних

4.3.1 Побудова моделей на основі даних ВВП

Вагомою характеристикою економічного росту держави є показники її валового внутрішнього продукту, тому далі ми розглянемо роботу моделей на статистичних даних ВВП високорозвиненої країни США, вибірка яких становить 415 елементів помісячно протягом 1985-2019 років.

Стан реальних статистичних даних за період 1985-2019 років на Рисунку 4.2:

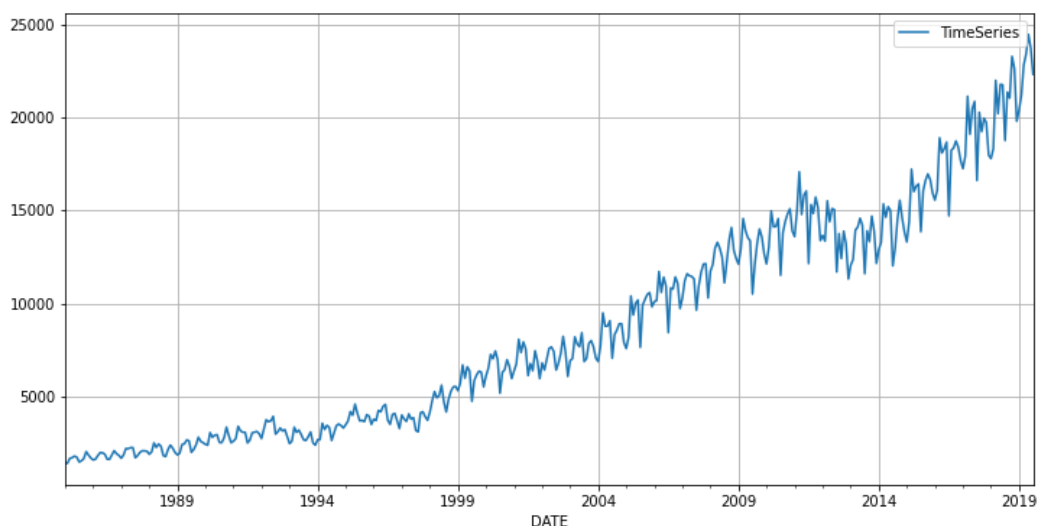


Рисунок 4.2 – Реальні дані ВВП США за період 1985-2019 років

Для того, щоб правильно побудувати модель з коректно обраним порядком моделі використовують автокореляційну функцію – АСФ(АКФ) та часткову автокореляційну функцію РАСФ(ЧАКФ), які зображені на наступних рисунках (Рисунок 4.3 та Рисунок 4.4 відповідно) :

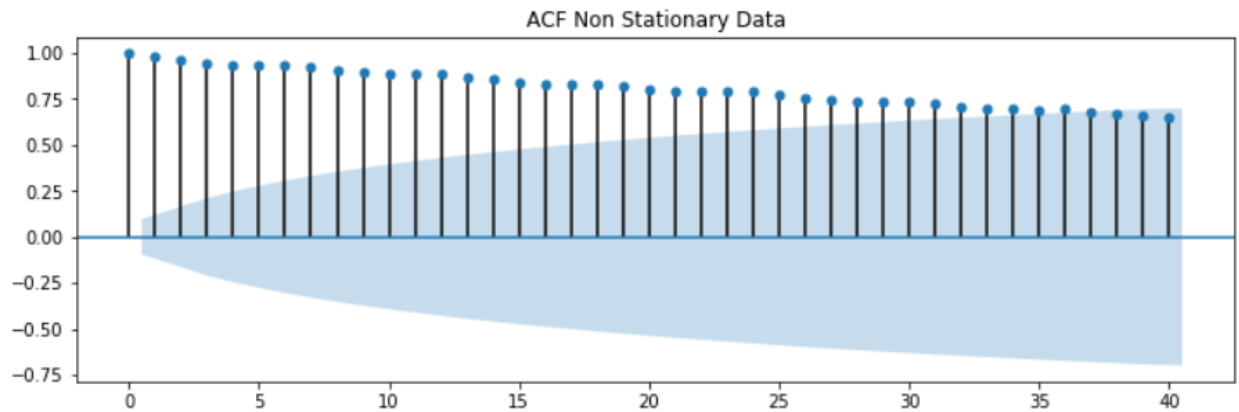


Рисунок 4.3 – АКФ для нестационарних даних ВВП США

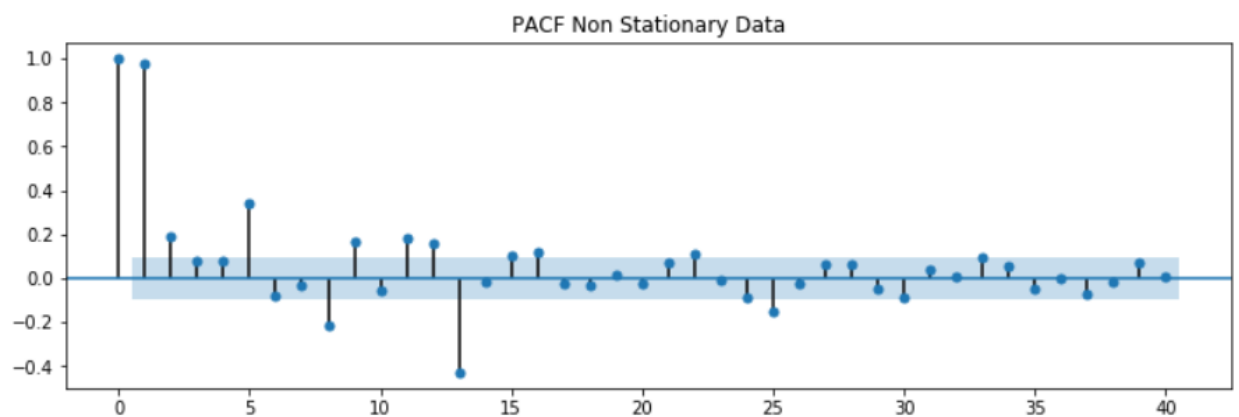


Рисунок 4.4 – ЧАКФ для нестационарних даних ВВП США

Автокореляційні функції були побудовані на 40 лагах, оскільки в ході експерименту, з'ясувалось, що це оптимальне число для того, щоб об'єктивно оцінити порядки моделей. За результатами побудови даних функцій на нестационарних даних та на стаціонарних даних було вирішено, що моделі можна будувати з параметром p , що відповідає першому, другому та тринадцятому порядку, адже вони значно виходять за виділену область. На основі стаціонарних даних на яких побудовані автокореляційні функції було обрано параметром $q = 2$.

Отже, оскільки порядки моделей підбрано можна виконувати побудову цих моделей та аналізувати адекватність.

Спочатку розглянемо авторегресійну модель першого порядку, що має наступні коефіцієнти:

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k) = 0.006723 + 0.191981 y(k-1) + \varepsilon(k)$$

Авторегресійне рівняння AR(2) має вигляд:

$$\begin{aligned} y(k) &= a_0 + a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k) = \\ &= 0.006700 + 0.238465 y(k-1) + 0.240346 y(k-2) + \varepsilon(k) \end{aligned}$$

Авторегресійне рівняння AR(13) має вигляд:

$$\begin{aligned} y(k) &= a_0 + a_1 y(k-1) + a_2 y(k-2) + a_3 y(k-3) + a_4 y(k-4) + \\ &+ a_5 y(k-5) + a_6 y(k-6) + a_7 y(k-7) + a_8 y(k-8) + a_9 y(k-9) + \\ &+ a_{10} y(k-10) + a_{11} y(k-11) + a_{12} y(k-12) + a_{13} y(k-13) + \varepsilon(k) = \\ &= 0.006617 + 0.424597 y(k-1) + 0.287398 y(k-2) + \\ &\quad + 0.124119 y(k-3) - \\ &\quad - 0.273453 y(k-4) - 0.005374 y(k-5) + 0.102960 y(k-6) + \\ &\quad + 0.035460 y(k-7) - 0.148099 y(k-8) + 0.092573 y(k-9) + \\ &\quad + 0.244186 y(k-10) + 0.102029 y(k-11) + 0.537296 y(k-12) + \\ &\quad - 0.199875 y(k-13) + \varepsilon(k) \end{aligned}$$

Візуалізуємо побудовані моделі AR(1), AR(2), AR(13) на порівняльному графіку на Рисунку 4.5.

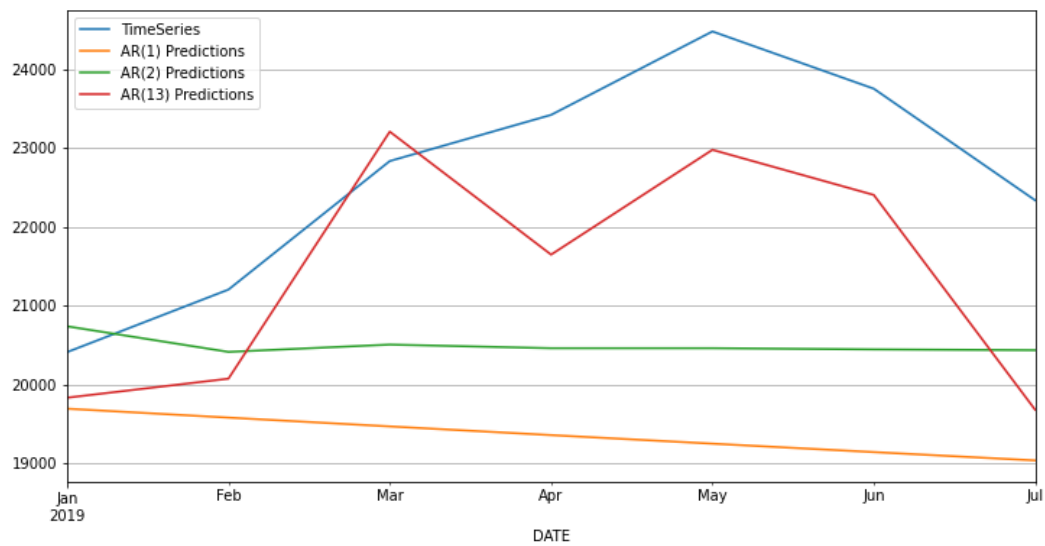


Рисунок 4.5 – Порівняння моделей $AR(1)$, $AR(2)$, $AR(13)$

На даному графіку можна побачити наочне порівняння роботи побудованих моделей з реальною вибіркою, тобто тестові дані, за період січень – липень 2019 року. Даний графік демонструє, що найкращу поведінку та динаміку ряду тестової вибірки ВВП США відтворює авторегресійна модель $AR(13)$.

Далі буде представлено графік, на якому виконана побудова моделі враховуючи тренди першого, другого та третього порядку в порівнянні з тестовою вибіркою на Рисунок 4.6.

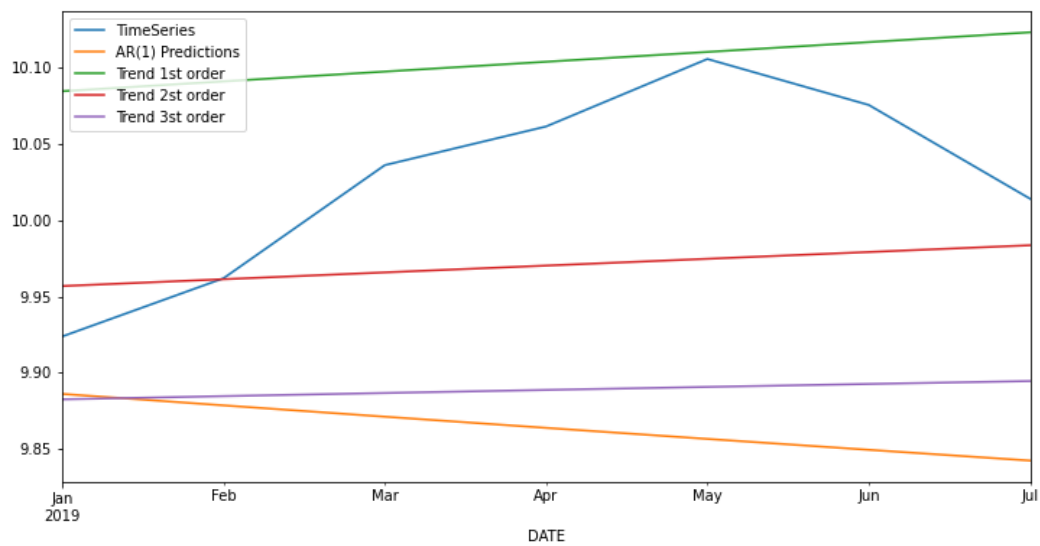


Рисунок 4.6 – Імплементация моделей з трендами

Авторегресійні моделі побудовані, тому перейдемо до побудови авторегресійних моделей з ковзним середнім, а саме ARMA(13,6):

$$\begin{aligned}
 y(k) = & a_0 + a_1 y(k-1) + a_2 y(k-2) + a_3 y(k-3) + a_4 y(k-4) + \\
 & + a_5 y(k-5) + a_6 y(k-6) + a_7 y(k-7) + a_8 y(k-8) + a_9 y(k-9) + \\
 & + a_{10} y(k-10) + a_{11} y(k-11) + a_{12} y(k-12) + a_{13} y(k-13) + \varepsilon(k) + \\
 & + \beta_1 \varepsilon(k-1) = \\
 = & 0.006617 + 0.424597 y(k-1) + 0.287398 y(k-2) + \\
 & + 0.124119 y(k-3) - 0.273453 y(k-4) - 0.005374 y(k-5) + \\
 & + 0.102960 y(k-6) + 0.035460 y(k-7) - 0.148099 y(k-8) + \\
 & + 0.092573 y(k-9) + 0.244186 y(k-10) + 0.102029 y(k-11) + \\
 & + 0.537296 y(k-12) - 0.199875 y(k-13) + \varepsilon(k) - 0.6443 \varepsilon(k-1)
 \end{aligned}$$

Зобразимо отриману модель на графіку та наведемо для порівняння тестову вибірку на проміжку січень – липень 2019 року (Рисунок 4.7).

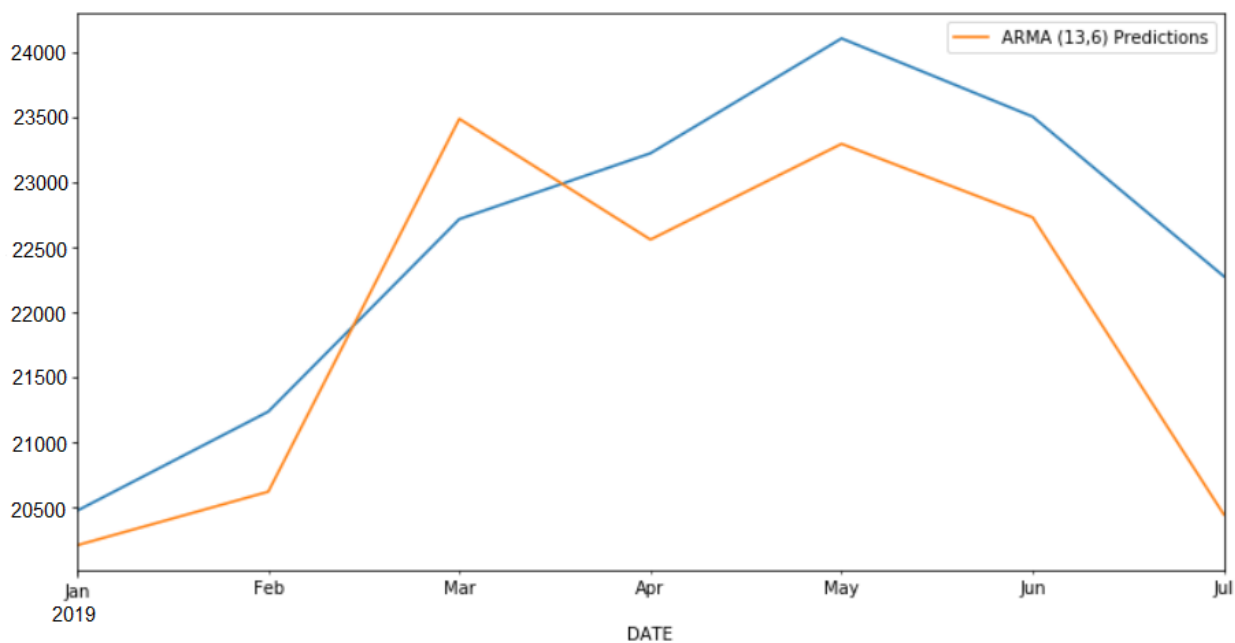


Рисунок 4.7 – Візуалізована модель ARMA(13,6)

Далі побудуємо модель ARIMA(13,1,2), що має порядок $d = 1$ на Рисунок 4.8

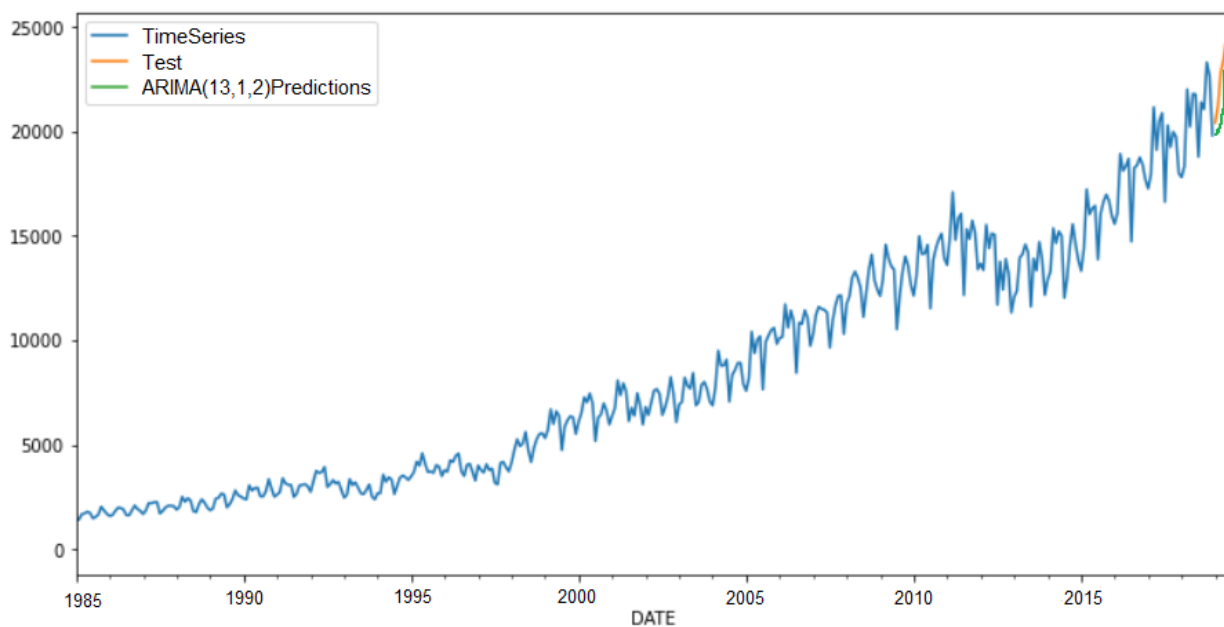


Рисунок 4.8 – Візуалізована модель ARIMA(13,1,2)

Отже, було побудовано основні моделі регресійного підходу прогнозування, а саме авторегресійні моделі різних порядків, авторегресійні моделі з ковзним середнім та авторегресійні моделі з ковзним середнім та сезонною зміною. Далі роботу цих моделей можна оцінити за допомогою рівня адекватності моделі та оцінок якості самого прогнозування, що і буде наведено нижче в порівняльній таблиці 4.1.

Таблиця 4.1 – Висновки по роботі побудованих моделей

Тип моделі	Критерії адекватності			Оцінки прогнозу		
	R^2	$\sum e^2(k)$	DW	$RMSE$	MAE	MAPE
$AR(1)$	0.6490	8.3109	0.8028	0.1420	0.1192	1.3594
$AR(13)$	0.9659	1.6883	2.3897	0.0651	0.0509	0.6144
$ARMA(13,6)$	0.9910	1.5164	2.2482	0.0617	0.0479	0.5917
$ARIMA(13,1,2)$	0.9922	1.3186	2.1294	0.0659	0.0524	0.5579

Аналізуючи результати моделювання та прогнозування з порівняльної таблиці, у якій наведені критерії адекватності та оцінки прогнозу побудованих моделей, можна зробити висновки, що найкращі результати має модель $ARIMA(13,1,2)$.

Отже, оскільки найкращі показники має модель $ARIMA(13,1,2)$, тепер побудуємо прогноз росту ВВП США на 12 місяців вперед, що зображений на Рисунку 4.9.

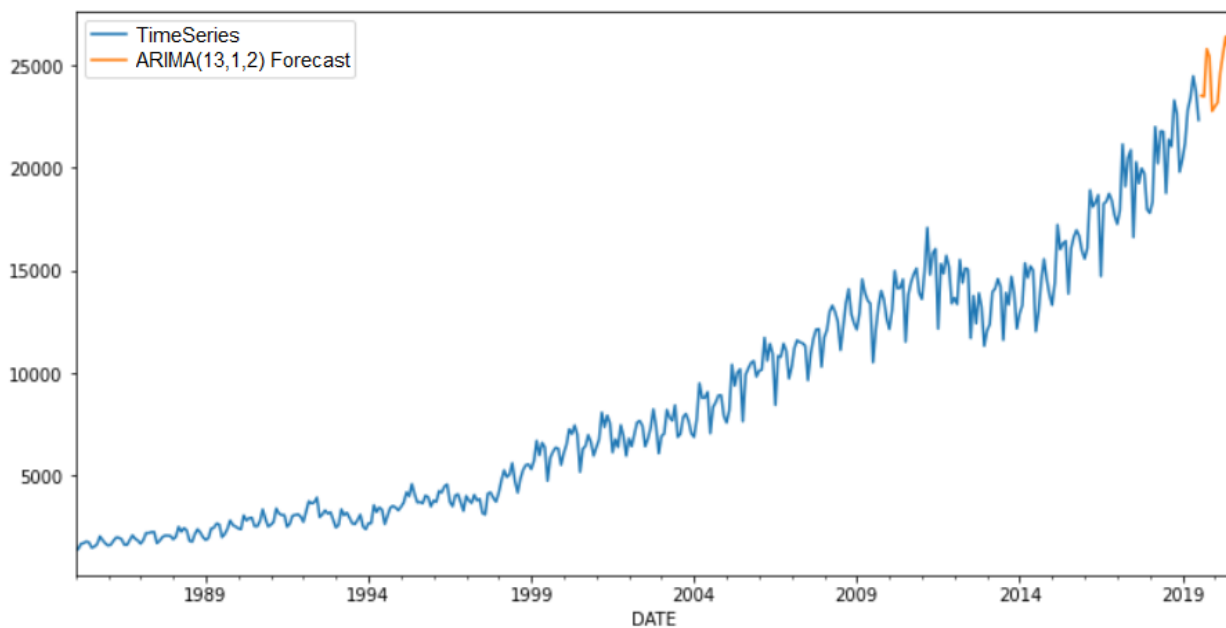


Рисунок 4.9 – Прогноз росту ВВП в США на 12 місяців

Тепер побудуємо нейронну мережу LSTM та навчимо її на тих самих даних ВВП США за період січень – липень 2019 року. Нейронна мережа містить чотири шари, де кожний шар має 50 нейронів при Dropout rate = 20%. Структуру моделі можна розглянути на Рисунку 4.10.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 1, 50)	10400
dropout (Dropout)	(None, 1, 50)	0
lstm_1 (LSTM)	(None, 1, 50)	20200
dropout_1 (Dropout)	(None, 1, 50)	0
lstm_2 (LSTM)	(None, 1, 50)	20200
dropout_2 (Dropout)	(None, 1, 50)	0
lstm_3 (LSTM)	(None, 50)	20200
dropout_3 (Dropout)	(None, 50)	0
dense (Dense)	(None, 1)	51
Total params: 71,051		
Trainable params: 71,051		
Non-trainable params: 0		

Рисунок 4.10 – Структура побудованої мережі

Функцією втрат в побудованій мережі є середня абсолютна похибка (MAE). Мережа має 100 епох з розміром батчів 64 та використовується оптимізатор Adam. Тренувальна вибірка містить 374 елементи та позначені синім кольором на графіку, а валідаційна вибірка має 41 елементів та позначена жовтим кольором, тобто розподілення має 90% та 10% відповідно. Прогнозуюча крива має зелений колір. А власне прогнозування розглянутих статистичних даних представлено на Рисунку 4.11.

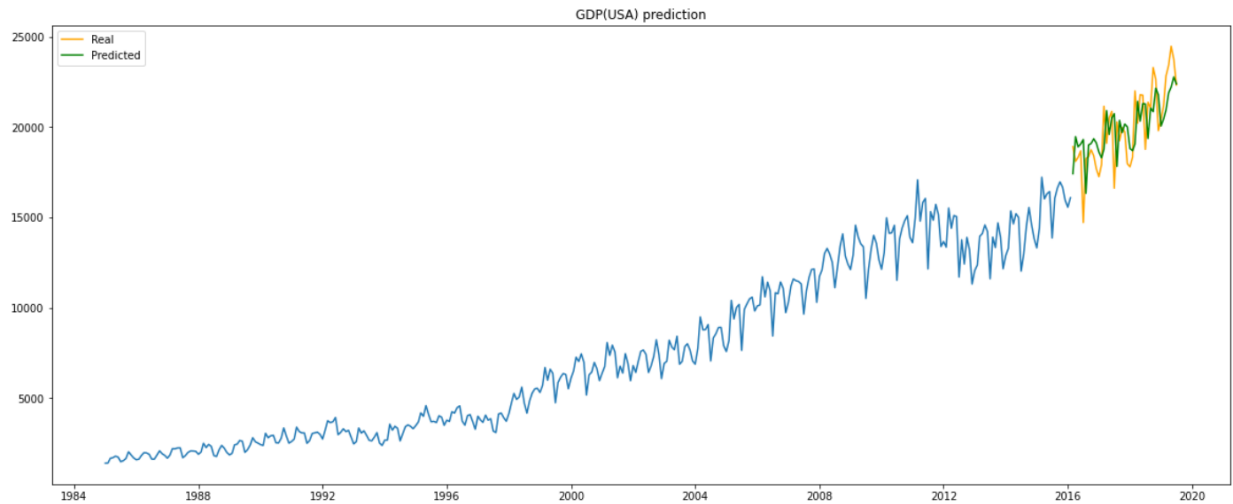


Рисунок 4.11 – Прогнозування нейронної мережі ВВП в США

Далі представлено порівняльну таблицю 4.2, на якій зображено результати моделювання $ARIMA(13,1,2)$ та нейронної мережі LSTM на основні вибірки ВВП США.

Таблиця 4.2 – Підсумки по роботі побудованих моделей

Тип моделі	Критерії адекватності			Оцінки прогнозу		
	R^2	$\sum e^2(k)$	DW	$RMSE$	MAE	MAPE
$ARIMA(13,1,2)$	0.9922	1.3186	2.1294	0.0659	0.0524	0.5579
$LSTM$	0.9189	0.5741	2.1387	0.4654	0.3126	0.5187

Можна бачити, що в обох моделей R^2 прямує до одиниці, сума квадратів похибок також прямує до нуля, а DW до двійки, що є показником ефективності використання обох моделей. Проте LSTM має дещо кращі показники. Те ж саме можна сказати про показники оцінки якості прогнозу, які в деякій мірі кращі в LSTM.

4.3.1 Побудова моделей на основі даних забруднення повітря

Вагомою характеристикою екологічного становища є показник забруднення повітря, тому далі ми розглянемо роботу моделей на статистичних даних забруднення повітря щодня, вибірка яких становить 365 елементів щоденно протягом 2014 року.

Стан реальних статистичних даних за період 2014 років на Рисунку 4.12:

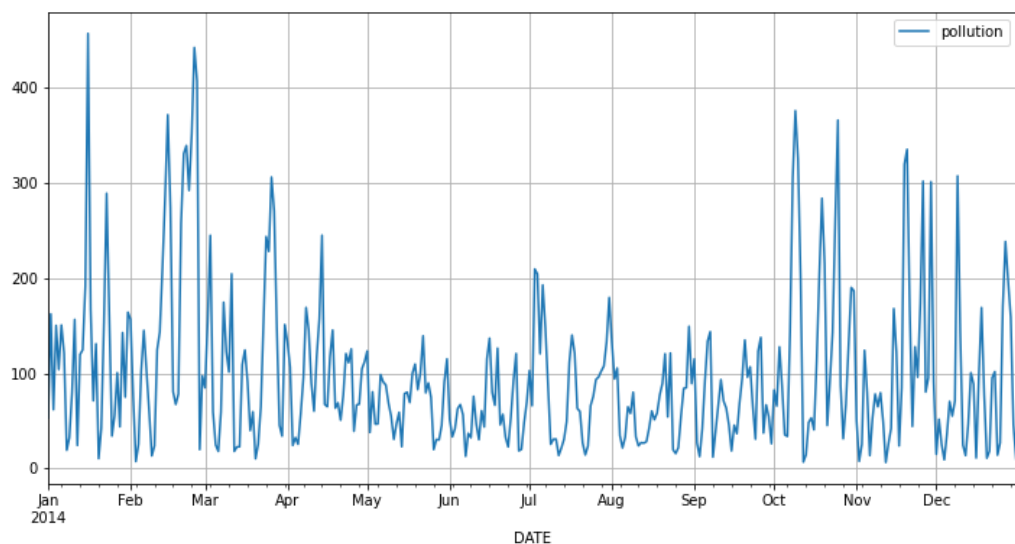


Рисунок 4.12 – Реальні дані забруднення повітря за період 2014 року

Для того, щоб правильно побудувати модель з коректно обраним порядком моделі використовують автокореляційну функцію – ACF(АКФ) та часткову автокореляційну функцію PACF(ЧАКФ), які зображені на наступних рисунках (Рисунок 4.13 та Рисунок 4.14 відповідно) :

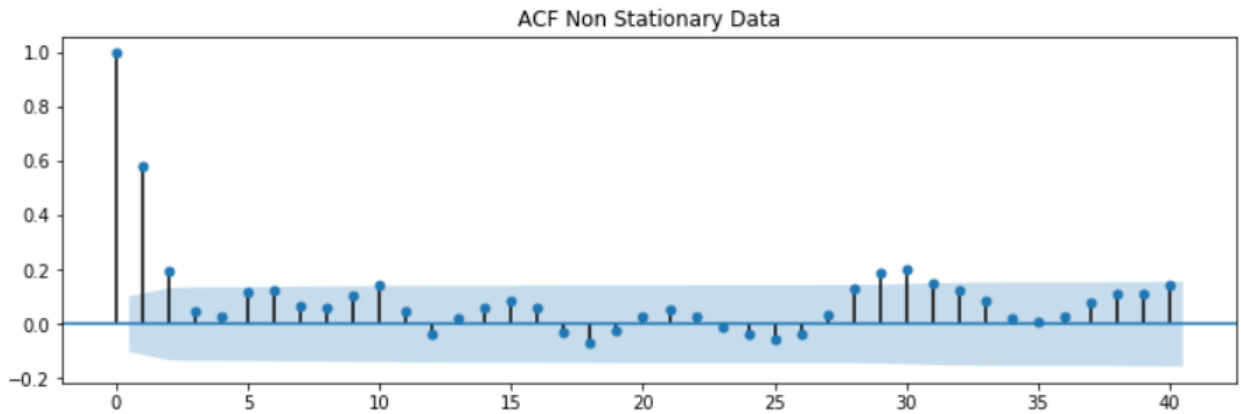


Рисунок 4.13 – АКФ для нестационарних даних забруднення повітря

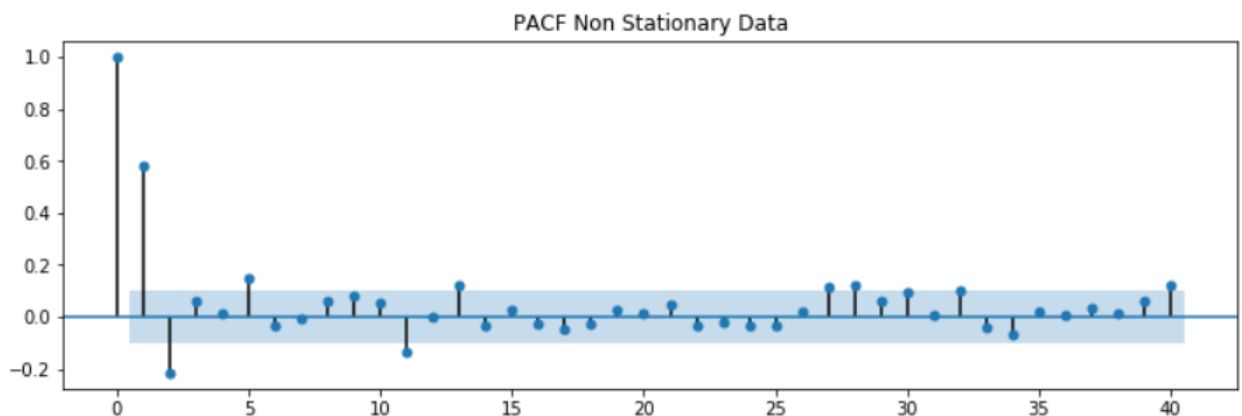


Рисунок 4.14 – ЧАКФ для нестационарних даних забруднення повітря

Автокореляційні функції були побудовані на 40 лагах, оскільки в ході експерименту, з'ясувалось, що це оптимальне число для того, щоб об'єктивно оцінити порядки моделей. За результатами побудови даних функцій на нестационарних даних та на стаціонарних даних було вирішено, що моделі можна будувати з параметром p , що відповідає першому та тринадцятому порядку, адже вони значно виходять за виділену область. На основі стаціонарних даних на яких побудовані автокореляційні функції було обрано параметром $q = 1$.

Отже, оскільки порядки моделей підібрано можна виконувати побудову цих моделей та аналізувати адекватність.

Спочатку розглянемо авторегресійну модель першого порядку, що має наступні коефіцієнти:

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k) = 3.7381 + 0.621858 y(k-1) + \varepsilon(k)$$

Авторегресійне рівняння AR(13) має вигляд:

$$\begin{aligned} y(k) = & a_0 + a_1 y(k-1) + a_2 y(k-2) + a_3 y(k-3) + a_4 y(k-4) + \\ & + a_5 y(k-5) + a_6 y(k-6) + a_7 y(k-7) + a_8 y(k-8) + a_9 y(k-9) + \\ & + a_{10} y(k-10) + a_{11} y(k-11) + a_{12} y(k-12) + a_{13} y(k-13) + \varepsilon(k) = \\ & = 0.006617 + 0.424597 y(k-1) + 0.287398 y(k-2) + \\ & + 0.124119 y(k-3) - 0.273453 y(k-4) - 0.005374 y(k-5) + \\ & + 0.102960 y(k-6) + 0.035460 y(k-7) - 0.148099 y(k-8) + \\ & + 0.092573 y(k-9) + 0.244186 y(k-10) + 0.102029 y(k-11) + \\ & + 0.537296 y(k-12) - 0.199875 y(k-13) + \varepsilon(k) \end{aligned}$$

Візуалізуємо побудовані моделі AR(1), AR(13) на порівняльному графіку на Рисунку 4.15.

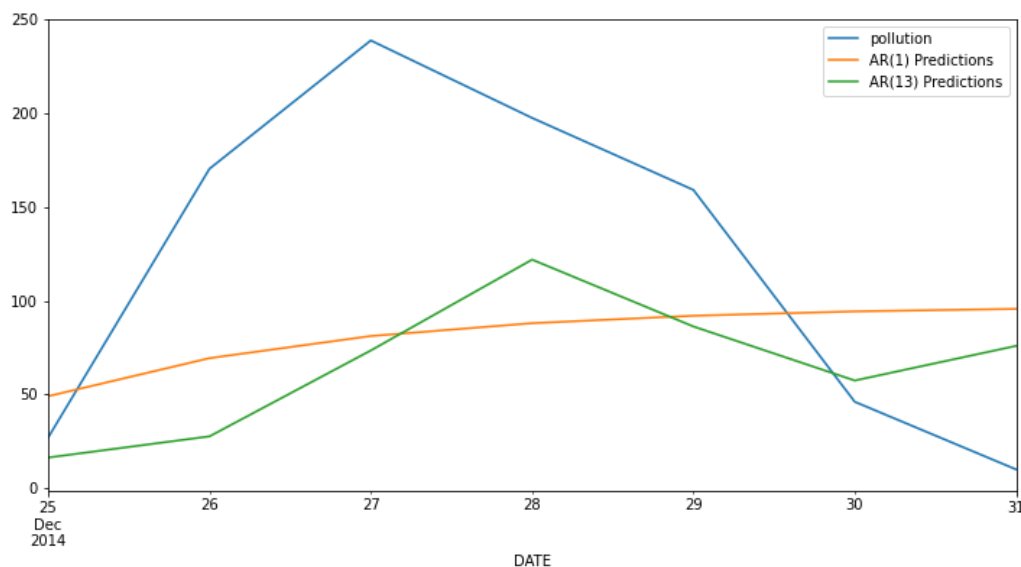


Рисунок 4.15 – Порівняння моделей AR(1), AR(13)

Далі побудуємо модель $ARIMA(13,1,1)$, що має порядок $d = 1$ на Рисунок 4.16

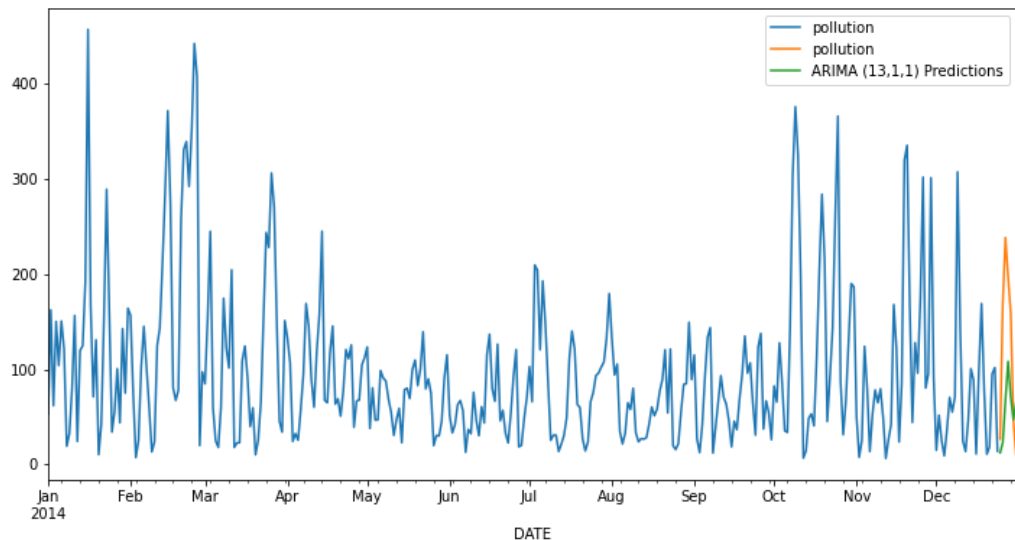


Рисунок 4.16 – Візуалізована модель $ARIMA(13,1,1)$

Отже, було побудовано основні моделі регресійного підходу прогнозування, а саме авторегресійні моделі різних порядків, авторегресійні моделі з ковзним середнім та сезонною зміною. Далі роботу цих моделей можна оцінити за допомогою рівня адекватності моделі та оцінок якості самого прогнозування, що і буде наведено нижче в порівняльній таблиці 4.3.

Таблиця 4.3 – Висновки по роботі побудованих моделей

Тип моделі	Критерії адекватності			Оцінки прогнозу		
	R^2	$\sum e^2(k)$	DW	$RMSE$	MAE	MAPE
$AR(1)$	0.6585	1.5421	0.9885	0.6532	0.4723	1.0155
$AR(13)$	0.6722	0.9255	2.2985	0.6486	0.4709	1.0056
$ARIMA(13,1,1)$	0.6932	0.7253	2.1352	0.5147	0.4012	0.6354

Аналізуючи результати моделювання та прогнозування з порівняльної таблиці, у якій наведені критерії адекватності та оцінки прогнозу побудованих моделей, можна зробити висновки, що найкращі результати має модель $ARIMA(13,1,1)$.

Отже, оскільки найкращі показники має модель $ARIMA(13,1,1)$, тепер побудуємо забруднення повітря на 7 днів вперед, що зображений на Рисунку 4.17.

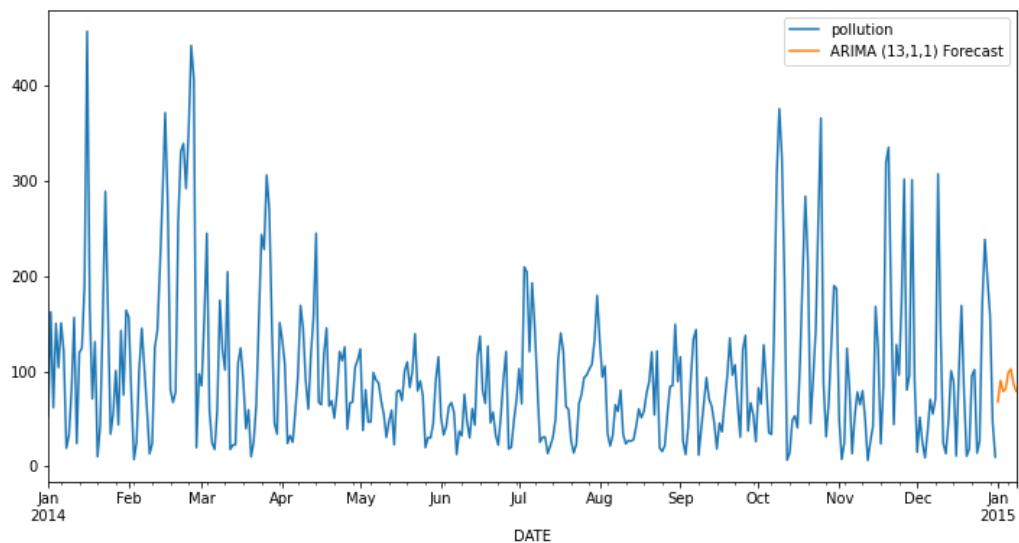


Рисунок 4.17 – Прогноз забруднення повітря на 7 днів

Тепер побудуємо нейронну мережу LSTM та навчимо її на тих самих даних забруднення повітря за період 2014 року.

Функцією втрат в побудованій мережі є середня абсолютна похибка (MAE). Мережа має 100 епох з розміром батчів 64 та використовується оптимізатор Adam. Тренувальна вибірка позначена синім кольором на графіку, а валідаційна вибірка позначена жовтим кольором, поділ вибірки виконаний наступним чином 75% та 25% відповідно. Прогнозуюча крива має зелений колір. А власне прогнозування розглянутих статистичних даних представлено на Рисунку 4.18.

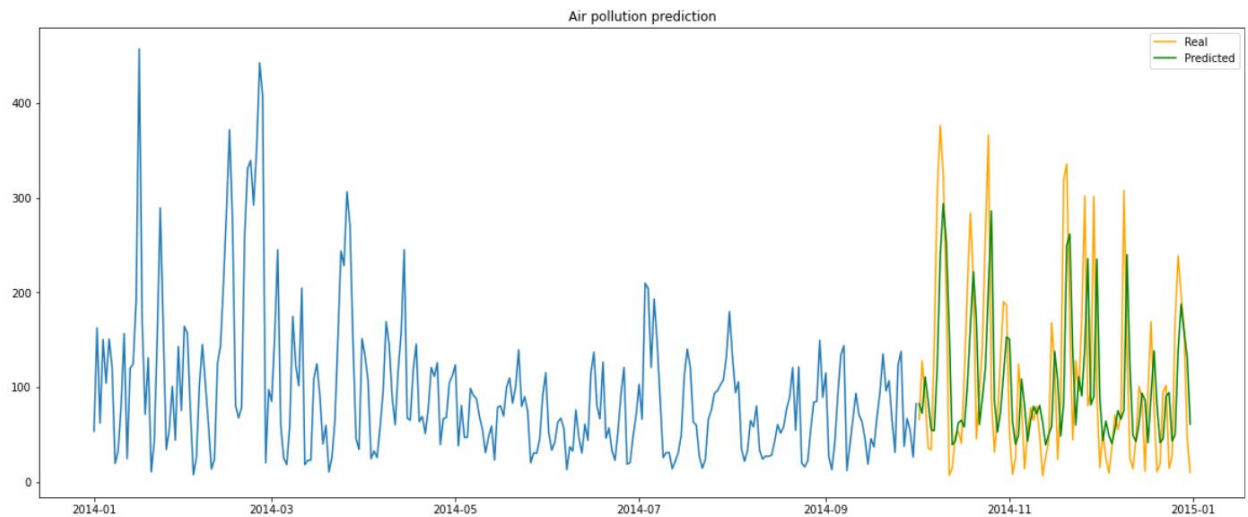


Рисунок 4.18 – Прогнозування нейронної мережі забруднення повітря

Далі представлено порівняльну таблицю 4.4, на якій зображено результати моделювання $ARIMA(13,1,1)$ та нейронної мережі LSTM на основні вибірки забруднення повітря.

Таблиця 4.4 – Підсумки по роботі побудованих моделей

Тип моделі	Критерії адекватності			Оцінки прогнозу		
	R^2	$\sum e^2(k)$	DW	$RMSE$	MAE	MAPE
$ARIMA(13,1,1)$	0.6932	0.7253	2.1352	0.5147	0.4012	0.6354
$LSTM$	0.7386	0.4215	2.0544	0.4163	0.2945	0.4354

Можна бачити, що в обох моделей R^2 прямує до одиниці, сума квадратів похибок також прямує до нуля, а DW до двійки, що є показником ефективності використання обох моделей. Проте LSTM має дещо кращі показники. Те ж саме можна сказати про показники оцінки якості прогнозу, які в деякій мірі кращі в LSTM.

Висновки до розділу 4

У даному розділі було продемонстровано побудовані моделі на основі регресійного підходу та інтелектуального аналізу даних за допомогою авторегресійних моделей різних порядків та рекурентної нейронної мережі разом з виконаним прогнозом на декілька кроків вперед. Також було практично реалізовано програмний продукт, що синтезує систему для моделювання та прогнозування нелінійних нестационарних часових рядів на реальних статистичних даних. У роботі моделі були апробовані на економічних даних – ріст ВВП США та екологічних даних – забруднення повітря. Система досить непогано аналізує динамічну зміну ряду, враховуючи сезонні збурення.

Було реалізовано регресійні моделі типу AR, ARMA, ARIMA та нейронну мережу LSTM. Результати моделювання перевірені відповідними критеріями адекватності моделей та в результаті чого було обрано найкращу модель. Коли найкращу модель вибрали та вибрали найкращу структуру нейронної мережі, виконується прогнозування навчених моделей. Результати прогнозування перевірялися за допомогою критеріїв оцінювання якості прогнозу.

За допомогою реалізованої системи користувач, який прагне аналізувати часові ряди на основі реальних статистичних даних та виконувати прогноз, може легко це зробити, адже всі етапи інтуїтивно зрозуміло сформовано та представлено в даній програмі.

РОЗДІЛ 5

РОЗРОБКА ВЛАСНОГО СТАРТАП ПРОЕКТУ

5.1 Сутність та особливості стартап проектів

Для створення стартап-проекту, перш за все, необхідно мати оригінальну ідею. Якщо ідеї ще немає, існують різні методи генерації інноваційних ідей. На основі використання морфологічного методу оберіть оптимальну ідею продукту.

Наший варіант виробу повинен виконувати функції інтуїтивно зрозумілого для кінцевого користувача інтерфейсного програмного забезпечення для моделювання та короткострокового прогнозування демографічних процесів з використанням авторегресійних методів, методу групового врахування аргументу та методу експоненційного згладжування. Усі основні параметри та проміжні рішення сформовані у таблиці 5.1.

Таблиця 5.1 – Морфологічна карта

Основні параметри	Проміжні рішення				
	1-ше	2-ге	3-є	4-те	5-те
Сфера застосування даних	Демографічні процеси	Фінансові процеси (курси валют, значення ВВП і т.д.)	Страхові випадки та кредити (ймовірні ризики та історія клієнта)	Медична галузь	Інше
Формат вхідних даних	Excel таблиця	Csv файл	База даних	Хмарне сховище	Інше
Методи прогнозування	Нейронна мережа	Байєсівська мережа	Регресійний аналіз	Методи Інтелектуального Аналізу Даних	Інше
Тип програми	Веб додаток	Локальна версія для ПК	Мобільний додаток	Програма у форматі VR	Інше
Мова програмування	Python	Eviews	SAS	Matlab	Інше
Тип ліцензій	Безкоштовна	Комерційна	Випробувальна версія	Студентська	Інше

Далі буде наведено план оновлення у вигляді таблиці 5.2

Таблиця 5.2 – План оновлення

№ з/п	Запитання	Відповідь
1	2	3
1	Частиною яких систем є продукт?	Аналітичні пакети програмних засобів
2	Які функції надсистеми може виконувати продукт? Як їх з ним пов'язати?	Прогноз макроекономічних показників у світі, динаміка демографічних процесів населення країни, кредитоспроможності клієнта банку
3	Чи можна розділити продукт на частини?	Базу даних можна зберігати локально або на загальному сервері, коли обчислювальні процеси виконуються у хмарі, а інтерфейс на девайсі користувача
4	Чи можна об'єднати (агрегувати) кілька елементів продукту в один?	Безперечно для комфортного та ефективного користування кінцевим продуктом буде краще зберігати всі елементи продукту в одному місці, для ефективного виконання обчислювальних процесів
5	Яким має бути ідеальний продукт?	Продукт має приймати на вхід файли будь-якого типу, а також враховувати специфікацію та особливості даних, зручний та інтуїтивно зрозумілий користувацький інтерфейс, низьковартісні обчислювальні ресурси
6	Що відбудеться, якщо вилучити цей продукт? Чим його можна замінити?	Готовий продукт на ринку аналітики такий, як Eviews, Deductor, ghjlerwsz SAS
7	Яким цей продукт був у минулому?	Eviews – незручний користувацький інтерфейс, відсутня можливість завантаження вхідних файлів певних типів, низька обчислювальна швидкість, не врахування особливостей нестационарних, нелінійних процесів та інші особливості
8	На розвиток яких функцій було спрямоване удосконалення продукту?	Сучасний користувацький інтерфейс, інші платформи використання, а також врахування особливостей процесів
9	Які функції залишилися «недорозвиненими»?	Додати композицію методів для більш точного прогнозування
10	Як можна зараз розвинути ці функції?	Доступне послідовне використання методів прогнозування, що веде за собою додаткові часові витрати та матеріальні

FutureViews – веб додаток для завантаження, аналізу та прогнозування нестационарних, нелінійних, динамічних процесів у різних сферах застосування. Використовуючи статистичні дані, які нам необхідно проаналізувати користувач їх підвантажує та застосовує різноманітні математичні операції для аналізу та прогнозування.

Етапи та продукти в послідовності заміщення можна розглянути в наступній таблиці 5.3

Таблиця 5.3 – Синхронізація завдань

Етапи	Продукти (послідовність заміщення)	
Назва	Збереження даних	Прогнозування на N кроків
Вчора	Локально файли	Обчислення з використанням регресійних моделей
Сьогодні	Бази даних, csv, Excel	Нейронні мережі
Завтра	Хмарне сховище	Штучний інтелект
Післязавтра	Персональне хмарне сховище із VR візуалізацією даних	Future View

5.2 Формування команди стартапу

Визначення важливості факторів щодо їх вкладу у створення та реалізацію стартапу наведені в таблиці 5.4

Таблиця 5.4 – Оцінка факторів реалізації стартапу

Фактор	Вага (важливість)
Ідея	6
Підготовка бізнес плану	7
Компетентність	8
Залученість і ризики	3
Обов'язки	5

Шкала оцінки (0-10)

Оцінювання особистого внеску кожного партнера у створення та реалізацію стартапу наведені в таблиці 5.5

Таблиця 5.5 – Оцінка факторів особистого внеску

Фактор	Гуць Євгеній	Тертичний Роман
Ідея	6	6
Підготовка бізнес плану	7	9
Компетентність	9	6
Залученість і ризику	2	4
Обов'язки	8	8

Шкала оцінки (0-10)

Визначення дольової участі у стартап проєкті кожного учасника зображені в таблиці 5.6.

Таблиця 5.6 – Оцінка дольової участі

Фактор	Вага (важливість)	Гуць Євгеній	Тертичний Роман
Ідея	6	6	6
Підготовка бізнес плану	7	7	9
Компетентність	8	9	6
Залученість і ризику	3	2	4
Обов'язки	5	8	8
Разом	402	203	199
Процент	100%	50,5%	49,5%

Шкала оцінки (0-10)

5.3 Розробка продукту для стартапу

Під час розробки продукту для стартапу вагоме місце займає розроблення MVP, меню якого зображене на рисунку 5.1

Рисунок 5.1 – Меню прогнозування.

1-й MVP: Меню прогнозування, де користувач завантажує статистичні дані у форматі .csv, за необхідністю перевіряє їх на стаціонарність за допомогою кнопки «Тест Діккі-Фуллера», за допомогою радіокнопки можна обрати модель і її порядок, та за потреби побудувати АКФ та ЧАКФ, вибравши власноруч кількість лагів. І, власне, зробити прогноз на визначену кількість років, натиснувши кнопку ОК.(Рисунок 5.2)

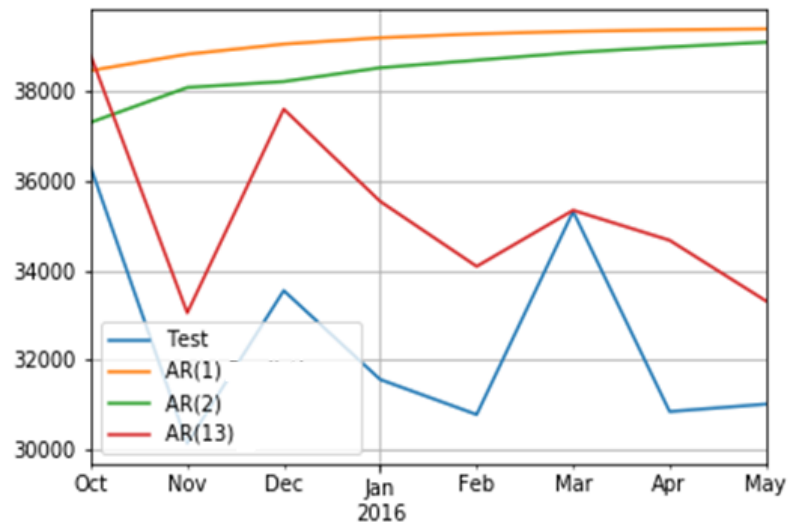


Рисунок 5.2 – Візуалізація побудованих моделей.

2-й MVP: Візуалізації побудови моделей для порівняння їхньої ефективності на реальних статистичних даних на часовому проміжку. Яку в подальшому можна використовувати для визначення оптимального порядку моделі авторегресійного прогнозування (Рисунок 5.3).

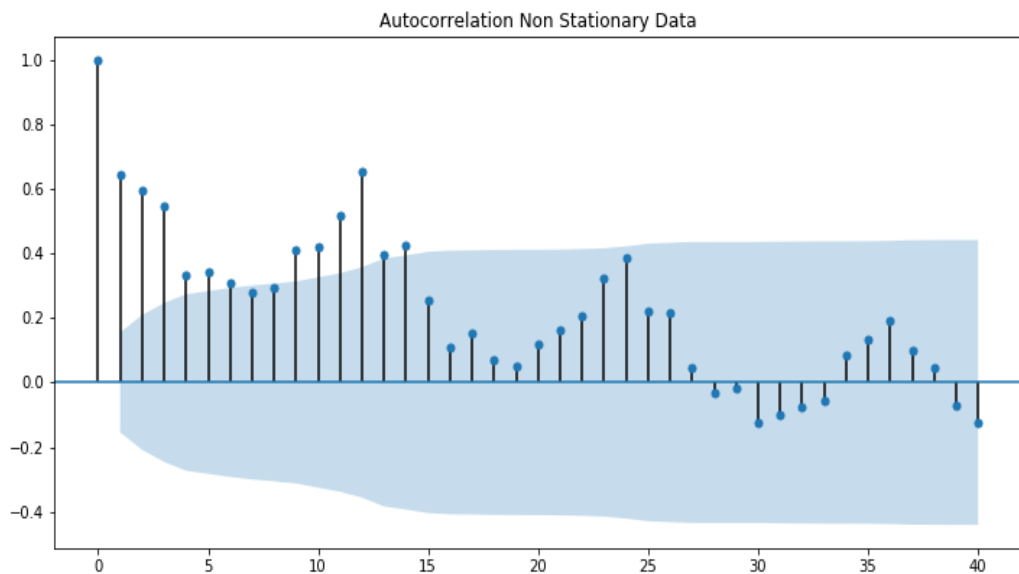


Рисунок 5.3 – Візуалізація методів Автокореляційної функції.

3-й MVP: Візуалізація методів Автокореляційної функції, а також часткової автокореляційної функції необхідної для обробки та вибору

порядку моделей при натисканні однієї з кнопок програмного продукту(Рисунок 5.4).

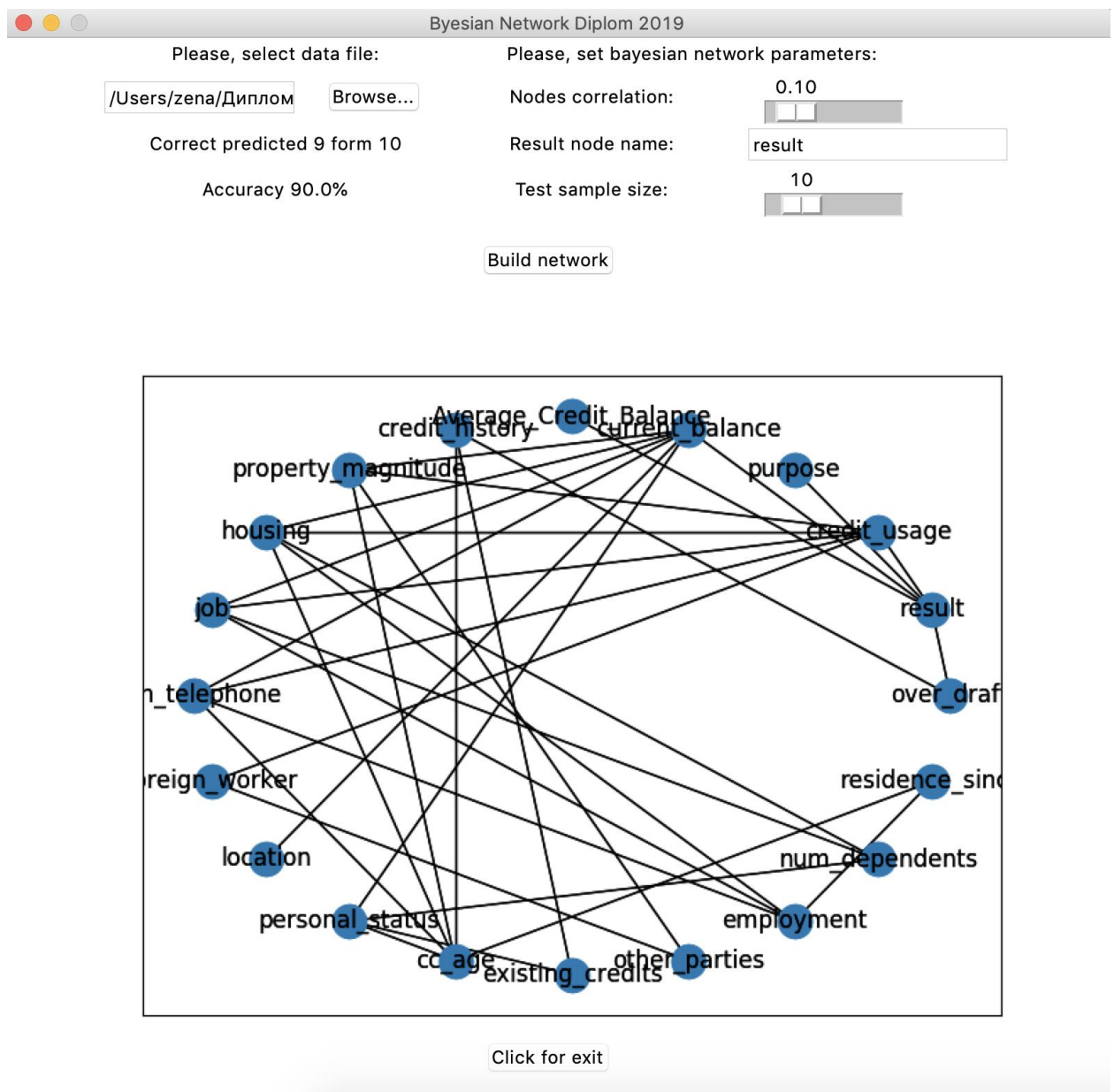


Рисунок 5.4 – Вікно побудови мереж Байєса.

4-й MVP: Застосування роботи мереж Байєса для побудови моделі, а також для прогнозу нелінійних, нестационарних процесів у фінансовій сфері, а саме аналізу кредитоспроможності позичальників в банках.

5.4 Розроблення бізнес-моделі стартапу

Бізнес-модель «Canvas» із всіма ключовими партнерами, видами діяльності, цілісною позицією, взаємовідносинами з клієнтами та споживчими сегментами цілісно відображено в таблиці 5.7.

Таблиця 5.7 – Бізнес-модель «Canvas»

Ключові партнери	Ключові види діяльності	Ціннісна пропозиція	Взаємовідносини з клієнтами	Споживчі сегменти
<ul style="list-style-type: none"> - державні служби статистики; - фондові біржі (NYSE, NASDAQ, D-J, ASX); - Финам фінансовий портал; - спеціалізовані пакети статистичних даних (Yahoo Finance; Google Finance; Federal Reserve Bank of St. Louis (FOREX)) 	<ul style="list-style-type: none"> - Розробка ПЗ - Вирішення проблеми актуальності статистичних даних та їхня повнота; - Робота відділу підтримки; 	<ul style="list-style-type: none"> - Актуальні статистичні дані та високоточний прогноз на будь-який проміжок часу; - Можливість безкоштовного випробувального терміну продукту; - Підтримка продукту протягом всього етапу його використання; - Зручний та інтуїтивно зрозумілий користувацький інтерфейс 	<ul style="list-style-type: none"> - Щоденна робота відділу Support 24/7 в телефонному режимі, а також за допомогою чат-ботів - Створення спеціалізованих спільнот - Автоматизоване обслуговування 	<ul style="list-style-type: none"> - Державні департаменти і різних сфер діяльності (фінансова, економічна, соціальна); - Міжнародні організації; - Підприємства, що потребують статистичні дані та прогноз зміни тенденцій, від яких залежить їхній бізнес
	Ключові ресурси <ul style="list-style-type: none"> - MySQL/MongoDB/Oracle - Android SDK - Microsoft 		Канали збуту <ul style="list-style-type: none"> - Реклама в соціальних мережах - Реклама на спеціалізованих платформах для ведення таргетного бізнесу - Особистий зв'язок та розсилка промо з державними установами 	

Далі буде наведено таблицю 5.8, що містить інформацію про структуру витрат та потоки надходження доходу

Таблиця 5.8 – Структура витрат та потоки надходження доходу

Структура витрат	Потоки надходження доходу
з переважною увагою до цінності	- Ліцензії програмного забезпечення;
- експлуатація серверів БД	- Продаж актуальних статистичних даних;
- ефект диверсифікації	- Технічна підтримка та обслуговування
- Залучення користувачів	продукту протягом всього циклу

Бізнес-модель «Lean Canvas» із всіма ключовими партнерами, видами діяльності, цілісною позицією, взаємовідносинами з клієнтами та споживчими сегментами цілісно відображено в таблиці 5.9.

Таблиця 5.9 – Бізнес-модель «Lean Canvas»

Проблема 1	Рішення 3	Унікальна торгова пропозиція 2	Прихована перевага 7	Споживчі 1
1) Відсутність безкоштовних платформ прогнозування	- Виористання альтернативних безкоштовних платформ обслуговування процесів	- Підтримка продукту протягом всього етапу його використання	- Сфокусованість на специфіку бізнесу	- Малий та середній бізнес в різних сферах
2) Відсутність підтримки протягом всього циклу співпраці	- Довгострокові договори та відділ підтримки	- Наявна гарантія точності прогнозу та актуальності даних, що базується на запатентованих аналітичних оцінках	- Унікальність методів вирішення проблеми	- Державні та міжнародні установи
3) Відсутність надання гарантії точності спрогнозованих процесів	- Запатентовані метрики оцінки прогнозу	- Зручний та інтуїтивно зрозумілий користувацький інтерфейс	- Регулярні бонуси за успішну співпрацю, що проявлятимуться згідно особливостей кожного партнера	- Дослідницькі та наукові центри

Продовження таблиці 5.9

<p>Ключові метрики</p> <p>6</p> <ul style="list-style-type: none"> - Оцінка повноти наданих статистичних даних та їх актуальність - Оцінка точності побудови моделі та її прогнозу, як короткострокового так і довгострокового -Стійкість моделі до впливу зовнішніх чинників в процесі 	<p>Канали збуту</p> <p>4</p> <p>Реклама в соціальних мережах</p> <ul style="list-style-type: none"> - Реклама на спеціалізованих платформах для ведення таргетного бізнесу - Особистий зв'язок та розсилка промо з державними установами
<p>Структура витрат</p> <p>5</p> <ul style="list-style-type: none"> - з переважною увагою до цінності <ul style="list-style-type: none"> - експлуатація серверів БД - ефект диверсифікації - Залучення користувачів 	<p>Потоки надходження доходу</p> <p>8</p> <ul style="list-style-type: none"> - Ліцензії програмного забезпечення; - Продаж актуальних статистичних даних; - Технічна підтримка та обслуговування продукту протягом всього циклу

5.5 Маркетингове планування стартапу

Характеристика потенційних клієнтів стартап-проекту та потреби, що формує ринок, його цільова аудиторія, відмінності в поведінці та вимоги споживачів до товару наведені в таблиці 5.10.

Таблиця 5.10 – Характеристика потенційних клієнтів стартап-проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1	ПЗ для швидкого та точного прогнозування як короткострокового так і довгострокового	- Державні департаменти; - Міжнародні організації; - Підприємства	Сфера діяльності та відповідно сфера застосування статистичних даних	-актуальність статистичних даних; -точність та якість прогнозу;
2	Зручний та інтуїтивно зрозумілий користувацький інтерфейс	- Державні департаменти; - Міжнародні організації; - Підприємства	1) Простота використання ПЗ 2)Високоточність результату	- Акцент на зручності та простоті інтерфейсу ПЗ - Якість та точність прогнозу

Визначення ринкових можливостей і загроз за вказаними параметрами оцінки зображено в порядку пріоритетності в таблиці 5.11.

Таблиця 5.11 – Ринкові можливості та загрози

Параметри оцінки	Можливості	Загрози
1. Конкуренція	Кращі результати прогнозування за короткий проміжок часу та низьку ринкову ціну	Можлива робота у від'ємну маржинальну дохідність та з кастомними клієнтами
2. Якість та точність прогнозу	Отримання високоточного результату прогнозу	Можлива велика ресурсозатратність в обчислювальних можливостях
3. Робота з клієнтами вузької сфери діяльності	Унікальність власної розробки, яка є дефіцитною на ринку	Перепрофілювання суто на кастомного клієнта, з чого слідує відсутність універсальності та довготривалість проекту

Ступеневий аналіз конкуренції на ринку представлений у вигляді згрупованої таблиці 5.12 всіх необхідних показників, що впливають на ті чи інші процеси, що нас цікавлять в процесі створення стартап проекту.

Таблиця 5.12 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Вказати тип конкуренції - Чиста	Можливість незалежної конкуренції між компаніями та встановлення самостійної ціни, що не контролюється ринковими цінами	Можливість прояву унікальності продукту та його інновація без встановлених стандартів
2. За рівнем конкурентної боротьби - Національна	Конкуренція на міжнародному ринку	Врахування особливостей локалізації клієнта, а не диктування своїх методів
3. За галузевою ознакою - Міжгалузева	Застосування ПЗ у різних сферах діяльності	Можливість легко інтегруватись у такі сфери як економіка, екологія, медицина, політика
4. Конкуренція за видами товарів: - Товарно-родова	Конкуренція між різноманітними видами товарів, що можуть виконувати схожі функції	Більш багатофактурна сфера застосування
5. За характером конкурентних переваг - Цінова/нецінова	Конкурентна перевага і цінова і нецінова. Якість продукту та її цінова категорія	Цікаві пропозиції для заохочення менш бюджетних компаній
6. За інтенсивністю - Не марочна	Спеціальної марки програмний продукт не має	Адаптація під інші виробництва призводить іноді до зміни функцій програми

Відомий маркетинголог М.Портер наводив канонічні параметри, за якими аналізувались галузі конкуренції, для обраного стартапу наведено та проаналізовано наступні, що зображені в таблиці 5.13.

Таблиця 5.13 – Аналіз конкуренції в галузі за М. Портером

	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари замітники
Складові аналізу	- HYDE; - NYSE; - Google Finance	- Статистика ООН; - Finam	HYDE LTD Corp., MADA ltd Comm	- Державні департамент; - Міжнародні організації; - Підприємства	Програмні продукти створені дослідницькими центрами
Висновки:	HYDE – вузькопрофільне середовище збереження даних; NYSE – не розглядає можливість прогнозування екологічних процесів Google Finance – високі затрати	Висока конкуренція на ринку, що має різноманітні шляхи розвитку	Фінансова залежність від спонсорів, що має сезонний характер та може впливати на роботоспроможність компанії	Зацікавлені в простоті використання та отриманні високоякісного результату	Дуже обмежений функціонал, що може впливати на якість результату

Далі в таблиці 5.13 наведено обґрунтування факторів конкурентноспроможності.

Таблиця 5.13 – Фактори конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
1	Унікальність користувацького інтерфейсу	На ринку дуже мало компаній випускають зручний та зрозумілий у використанні інтерфейс, щоб персонал без додаткових туторіалів та навчальних лекцій не інтегрував інтерфейс в робочий процес
2	Математична та статистична база	Використання найновіших методик розрахунку прогнозу з урахуванням особливостей різноманітних процесів
3	Актуальність даних	Використання актуальних статистичних даних, що оновлюються якнайшвидше, залежно від характеру процесів та специфіки бажаного результату

Нарешті можна продемонструвати SWOT- аналіз стартап-проекту, що зображений в таблиці 5.14.

Таблиця 5.14 – SWOT- аналіз стартап-проекту

<p>Сильні сторони:</p> <ul style="list-style-type: none"> - Унікальні підходи до математичного розрахунку; - Високоякісні технології; - Великий спектр застосування; - Інтуїтивно зрозумілий інтерфейс 	<p>Слабкі сторони:</p> <ul style="list-style-type: none"> - Молода компанія на ринку; - Відсутність налагоджених зв'язків з клієнтом; - Відсутність початкового фінансування;
<p>Можливості:</p> <ul style="list-style-type: none"> - Заохочення малого та середнього бізнесу; - Робота висококваліфікованих спеціалістів; - Інтегрування процесів прогнозування в інші сфери діяльності; - Іноваційні підходи та цікаві пропозиції 	<p>Загрози:</p> <ul style="list-style-type: none"> - Можливість поглинання конкурентами; - Неактуальність на локальному ринку; - Втрата стартової компенсації

5.6 Бізнес-план стартап проекту

Для хорошого бізнес- плану стартап проекту дуже важливо мати визначення базової стратегії розвитку, що підготована та сформована в таблиці 5.15.

Таблиця 5.15 – Базова стратегія розвитку

№ п/п	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
1	Промо-розсилка в зацікавлені державні департамент та дослідні центри, а також пропозиції малому та середньому бізнесу.	Випробувальні та безкоштовні версії ПЗ на обмежений час та для кастомних клієнтів. Цікаві пропозиції в напрямку підтримки власного продукту на стороні клієнта	Максимальна рекламна компанія та охопат цільової аудиторії. Запровадження ПЗ та демонстрація роботи. Підтримка та оновлення ПЗ протягом всього життєвого циклу.

А також не менш важливо мати вже сформовану таблицю 5.16, що містить визначення базової стратегії конкурентної поведінки

Таблиця 5.16 – Базова стратегія конкурентної поведінки

№ п/п	Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки*
1	Ні, але має унікальні ідеї	Орієнтована на пошук нових партнерів, шляхом зацікавлення в новому підході до ведення їхнього бізнесу	Основні характеристики будуть незмінні, оскільки вони базуються на математичних законах, а підходи до рішення будуть змінюватись	Стратегією будуть особливі пропозиції та інноваційні підходи

Визначення ключових переваг концепції потенційного товару займає також велике значення в аналізі розвитку стартапу, тому для ефективного виконання цього завдання створена таблиця 5.17.

Таблиця 5.17 – Ключові переваги концепції товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Отримання актуальних даних	Високоточні статистичні дані в режимі реального часу	Оновлення даних якнайшвидше в режимі реального часу залежно від специфіки даних та поставлених бізнес-цілей
2	Високоякісний довгостроковий/короткостроковий прогноз	Побудований прогноз, який відповідає усім критерія оцінки	Використання найновіших методик розрахунку прогнозу з урахуванням особливостей різноманітних процесів та поставленого бізнес-плану користувача
3	Інтуїтивно зрозумілий інтерфейс	Простота та лаконічність зовнішньої оболонки ПЗ, яку бачить та використовує користувач	На ринку дуже мало компаній випускають зручний та зрозумілий у використанні інтерфейс, щоб персонал без додаткових туторіалів та навчальних лекцій не інтегрував інтерфейс в робочий процес

Опис всіх рівнів моделі товару, який буде результатом стартапу наведений в таблиці 5.18.

Таблиця 5.18 – Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові
I. Товар за задумом	Сервіс прогнозування
II. Товар у реальному виконанні	Сервіс надання актуальної статистичної інформації та побудова на основі неї високоякісного короткострокового/довгострокового прогнозу враховуючи особливості обраних динамічних процесів; Сучасний зрозумілий інтерфейс для візуалізації та запуску усіх обраних процесів
III. Товар із підкріпленням	Можливість автономного виконання прогнозування на вказаний період часу та тригерні запуски функціоналу залежно від особливостей поведінки нестационарних нелінійних процесів у різноманітних сферах діяльності

Формування системи збуту наведено в таблиці 5.19.

Таблиця 5.19 – Система збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
1	Державні організації міжнародного рівня	Виняткові права на постачання товару з відсутністю посередників та дотримання норм чинного законодавства	канал нульового рівня	Розсилка промо в держоргани відповідальним особам
2	Малий та середній бізнес	Виняткові права на постачання товару з відсутністю посередників та дотримання комерційної таємниці про співробітництво	канал нульового рівня	Розсилка промо на офіційну робочу електронну пошту

Концепція маркетингових комунікацій можна зобразити у форматі таблиці 5.20, що представлена нижче.

Таблиця 5.20 – Маркетингові комунікації

№ п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Пошук безкоштовних або низьковартісних варіантів продукту	Спеціалізовані групи та канали зв'язку в мережі інтернет	Дешевизна	Запропонувати вигідну та цікаву пропозицію з подальшою співпрацею	Реклама, що містить демонстративний характер особливостей нашого продукту
2	Шляхом укладання взаємовигідних угод/договорів з двостороннім обміном	Державні канали зв'язку з потенційними партнерами	Надійна та довготривала співпраця	Представити основні переваги над конкурентами	Офіційне звернення у вигляді листа, що містить основні пункти співпраці

5.7 Патентний пошук

Необхідно дослідити ринок вже існуючих патентів, що можуть відповідати створюваному стартап проекту. Тому деякі з них наведені в таблиці 5.21.

Таблиця 5.21 – Результати патентного пошуку для ідеї стартап-проекту

№	Країна	Шифр документа	ПІБ автора ідеї	Назва документа і короткий опис ідеї
1	USA	US2016110812A1	Mun Johnathan	“Project economics analysis tool” - The present invention is in the field of corporate finance, corporate capital investments, economics, math, risk analysis, simulation, decision analysis, and business statistics, and relates to the modeling and valuation of investment decisions under uncertainty and risk within all companies, allowing these firms to properly identify, assess, quantify, value, diversify, and hedge their corporate capital investment decisions and their associated risks.
2	Canada	CN201710156129A	Electric Power Res Institute State Grid Jibei Electric Power Company Limited	“Method and device for predicting number of customer complaints” - The present invention relates to data processing technology, in particular to a method and a device for predicting the number of customer complaints.

Продовження таблиці 5.21

№	Країна	Шифр документа	ПІБ автора ідеї	Назва документа і короткий опис ідеї
3	USA	US2008002389W	Mordecai David K A	“system and method for dynamic path- and state-dependent stochastic control allocation” - The present invention relates to transactions, and, more particularly, to mitigating risk and optimizing gain between multiple parties in multiple transactions in involving state-dependent and path-dependent conditions where outcomes maybe jointly interdependent between parties and across transactions.
4	USA	US20141421112A	Mun Johnathan	“Qualitative and quantitative modeling of enterprise risk management and risk registers” - The present invention is in the field of corporate finance, corporate capital investments, economics, math, risk analysis, simulation, decision analysis, and business statistics, and relates to the modeling and valuation of investment decisions under uncertainty and risk within all companies, allowing these firms to properly identify, assess, quantify, value, diversify, and hedge their corporate capital investment decisions and their associated risks. Specifically, the present invention looks at starting from a comprehensive qualitative risk register and moving the analysis into the realms of quantitative risk modeling, simulation, and optimization.

5.8 Правові аспекти реалізації стартапів, інтелектуальна власність та патентування

1. Резюме

Найменування проекту: **FutureViews**;

Характеристика організації:

- ТОВ “Ф’ючер В’юз”;
- товариство з обмеженою відповідальністю;
- кількість співробітників: 2;
- контактні дані: futureviews_ltd@gmail.com, +380978987215;
- Тертичний Роман Віталійович, 22, Head of Department Analytics.

FutureViews - “start-up” проект, реалізований як мобільний додаток для Android та IOS, а також веб-сайт для створення власних короткострокових або ж довгострокових динамічних процесів базованих на обраних статистичних даних.

Персонал:

1) Тертичний Роман Віталійович — Data Analytic;

2) Гуць Євгеній Віталійович - Web Developer.

- На сьогоднішній день існують сервіси для обробки статистичних даних та створення короткострокових або ж довгострокових динамічних процесів, проте вони є або вузьконаправленими тільки на певної сфери застосування або ж їх використання є багатозатратним та немає підтримки сервісу на постійній основі.
- Використання сучасних технологій для обробки актуальних статистичних даних в будь-якій сфері діяльності.
- Власні ресурси відсутні.
- Вихід на міжнародний рівень після двох років роботи, вихід на перше місце серед конкурентів.
- Потреба інвестиції у серверну частину проекту, інвестиції будуть повертатися за допомогою преміум версії підписки на сервіс, а також реклами.
- Використання ліцензій для випробувальних термінів для поширення технологій в наукових центрах та лабораторіях.
- Застосування хмарних середовищ Google для зберігання та обробки статистичної інформації з подальшим впровадженням в технологію.
- Половина середнього та великого бізнесу використовують статистичні дані для планування своєї стратегії та розподілення ресурсів.
- Самостійне створення необхідних сервісів, які є аналогами Google Analytics та поступовий перехід виключно на власний сайт.

2. Опис підприємства

Команда складається з двох осіб розробників, та має комерційний досвід у веб-технологіях, а саме PHP, Angular JS, React JS; розробці мобільних додатків (Kotlin, Swift), і роботи з базами даних Oracle, MySQL.

3. Опис продукту

FutureViews – веб додаток для завантаження, аналізу та прогнозування нестационарних, нелінійних, динамічних процесів у різних сферах застосування. Використовуючи статистичні дані, які нам необхідно проаналізувати користувач їх підвантажує та застосовує різноманітні математичні операції для аналізу та прогнозування.

Розкажіть про асортимент вашої продукції, перерахуйте основні найменування ваших товарів і послуг, якщо їх занадто багато, об'єднайте в групи

Наш продукт - це комплект веб-сервіс (futuresviews.com) + мобільний додаток Android + мобільний додаток iOS.

Опишіть функціональне призначення продукції, для яких цілей вона призначена. Які потреби задовольняє, які проблеми допомагає вирішити.

Оскільки наш сайт має два глобальні підходи до використання даних та отримання бажаного результату з мінімальними затратами, користувач має свободу вибору. Перше - це зручний веб-сервіс, який поєднує в собі хмарне середовище, де можна знайти безліч корисного(найактуальніші статистичні дані структуризовані на вибір користувача відповідно його вимог з дуже зручною панеллю пошуку та голосовим асистентом), а також додати власноруч власні дані, які адаптуються до необхідної моделі. Перевага сервісу в тому, що користувачеві не потрібно зберігати всі етапи обробки та проміжні результати на власному пристрої, для цього створюється власний

кабінет, де його дані будуть надійно захищені відповідно до всесвітньої технології Security Personal Data Storage. Друге – це можливість швидкого доступу до Ваших даних через мобільні додатки, що реалізовані як на Android та і на iOS. Інтуїтивно зрозумілий інтерфейс та можливість прив'язки до власного кабінету користувача спростить Вашу роботу та допоможе зберегти дорогоцінні хвилини в доступу до актуальної інформації по бізнесу.

Уявіть основні технічні характеристики вашого продукту (Не перевантажуйте читача спеціалізованими термінами).

Отже, наш веб-сервіс створений за допомогою Angular JS та в основі має одну з найпопулярніших та найнадійніших нерелятивних баз даних Oracle, основні методи та моделі високоякісного прогнозу реалізовані на Python. Та мають онлайн доступ до швидкісного оновлення даних, що прив'язані до Загальної Світової Статистичної Бази (ЗССБ).

Наведіть приклади використання продукції.

- вибір статистичних даних із категорій запропонованих веб-сервісом;
- завантаження власних статистичних даних;
- вибір адаптації статистичних даних ;
- вибір параметрів для виконання прогнозу;
- вибір типу оцінки якості моделювання та прогнозування;
- перерахуйте основні етапи виробництва вашого продукту;
- побудова бази даних;
- розробка веб-сервісу;
- розробка додатку Android/ iOS;
- розробка хмарного сховища;
- підтримка та розробка нового функціоналу на всіх платформах.

Наскільки ваш продукт є універсальним або унікальним для кожного клієнта, в чому це проявляється.

Будь-який користувач може налаштувати інтерфейс на свій лад. Вибір платформи для роботи з програмою, яка інтегрується з усіма пристроями за допомогою власного кабінету. Кожен користувач може налаштувати продукт відповідно до свого типу бізнесу та особливостями ведення його відповідно географічного розташування.

Наскільки ваші товари або послуги відповідають прийнятим стандартам

Дана продукція повністю відповідає вимогам законодавства, менталітету та особливостям ведення тамтешнього бізнесу країни, де застосовується наш продукт. Торгової марки наразі немає. Дизайн сайту та додатків створений для надшвидкого та інтуїтивного використання. Продукт є унікальним не лише в Україні, а також у решті світу. Він є інтуїтивно простим у використанні, людина будь-якого сфери діяльності користуватися всіма функціями без особливої підготовки.

Товар та послуги повністю відповідають стандартам.

На якій стадії знаходиться продукт в даний час (ідея, робочий проект, дослідний зразок, серійне виробництво і т.д.)

Продукт знаходиться на стадії розробки додатку.

Вкажіть, чи є можливості для подальшого розвитку продукту

Планується перевести продукт на вищий рівень реалізації в плані покращення хмарного середовища з елементами автоматизації за допомогою Штучного інтелекту для полегшення роботи.

Проаналізуйте продукцію конкурентів, яка існує на ринку, відповідно до наступних критеріїв

На даний момент, конкурентами на нашому ринку організації, що займаються прогнозуванням вузьких сфер діяльності, проте їх достатня кількість та вони мають налагоджені зв'язки збуту продукції та вже вкладені договори про співпрацю.

Що відзначають споживачі в своїх відгуках про ваш продукт. Якщо є письмові відгуки, або рекомендації, приведіть їх в додатку.

Сьогодні продукт має схвальні відгуки на рівні університетських установ та викладачів, які допомагали з рецензією.

4. Аналіз ринку

Аналіз ринку вимагає табличного представлення pest – аналізу, що наведено в таблиці 5.22.

Таблиця 5.22 – Pest - аналіз

Групи чинників	Події/ чинники	Небезпека/можливість	Вірогідність події або прояву чинника	Важливість чинника або події	Вплив на підприємство	Програма дій
Політичні	1 Державний вплив у галузь	2	1	1	1	Робота з державними сервісами
	2 Зміна законів України	1	1	2	2	Переобладнання роботи
Економічні	1 Ринок і торговельні цикли	3	2	3	3	Дослідження ринку
	2 Витрати на підтримку	1	1	2	3	Дослідження фінансової частини
Соціальні	1 Вплив ЗМІ	1	1	1	1	Робота з ЗМІ
	2 Демографічні зміни	2	2	3	3	Реклама, та робота з ЗМІ
Технологічні	1 Нові продукти	2	3	2	3	Впровадження нових технологій, цінова політика
	2 Розвиток технологій	2	3	3	2	Впровадження нових технологій

Сильні та слабкі сторони проекту наведені в таблиці 5.23 нижче.

Таблиця 5.23 – SWOT – аналіз стартап-проекту

<p>Сильні сторони:</p> <ul style="list-style-type: none"> - Унікальні підходи до математичного розрахунку; - Високоякісні технології; - Великий спектр застосування; - Інтуїтивно зрозумілий інтерфейс 	<p>Слабкі сторони:</p> <ul style="list-style-type: none"> - Молода компанія на ринку; - Відсутність налагоджених зв'язків з клієнтом; - Відсутність початкового фінансування;
<p>Можливості:</p> <ul style="list-style-type: none"> - Заохочення малого та середнього бізнесу; - Робота висококваліфікованих спеціалістів; - Інтегрування процесів прогнозування в інші сфери діяльності; - Іноваційні підходи та цікаві пропозиції 	<p>Загрози:</p> <ul style="list-style-type: none"> - Можливість поглинання конкурентами; - Неактуальність на локальному ринку; - Втрата стартової компенсації

Матриця оцінки загальноєкономічних факторів впливу зовнішнього середовища на реалізацію стартап-проекту наведена в таблиці 5.24.

Таблиця 5.24 – Матриця оцінки факторів

Групи та окремі фактори		Оцінка залежності бізнесу від впливу зовнішніх факторів*					Сприятливі можливості	Загрози та небезпеки
Макроекономічні фактори	Посилення інфляції		2				Немає	Немає
	Зростання доходів населення та підвищення купівельної спроможності			3			Збільшення цін	Конкуренція
	Посилення диференціації доходів населення		2				Немає	Немає
	Зниження процентних ставок на банківський кредит		2				Немає	Немає
	Нестабільність макроекономічної ситуації			3			Немає	Немає
	Зростання економіки і розширення місткості ринку				4		Збільшення клієнтів і цін	Конкуренція
Елементи конкурентного середовища	Ринкові фактори				4		Збільшення цін	Конкуренція
	Вплив конкурентів					5	Впровадження нових технологій	Втрата клієнтів
	Вплив споживачів				4		Немає	Конкуренція
	Вплив постачальників		2				Немає	Немає
Соціально-демографічні характеристики	Зміни смаків споживачів, посилення рівня диференціації споживчих запитів та поява вільних ринкових сегментів, ін.			3			Отримання нових користувачів	Втрата користувачів та партнерів
Зміни в системі державного регулювання	Зміни в оподаткуванні		2				Немає	Немає
	Запровадження ліцензування та обов'язкової стандартизації продукції і послуг			3			Втрата деяких конкурентів	Втрата партнерів
	Посилення державного контролю та вимог			3			Немає	Втрата партнерів
	Посилення корупції	1					Немає	Немає
	Регулювання ціноутворення				4		Немає	Втрата користувачів
	Зміни в державних замовленнях	1					Немає	Немає
Галузеві фактори	Посилення конкуренції					5	Немає	Втрата клієнтів

5. План збуту та маркетингу

Планується два види отримання прибутку від створеного проекту. Перший з них це отримання коштів за куплені ліцензії різних типів, серед яких є й безкоштовні ліцензії, що покликані на заохочення до співпраці науково-дослідницьких центрів та перспективних студентів технічних вишів, а також безпосередньо категорія Преміум ліцензії, які і становитимуть вагому нішу нашого доходу. І другим видом доходу, буде монетизації кількості скачувань додатків на пристрої через Google Play та AppStore.

Вартість продукції буде розраховано з урахуванням витрат на підтримку серверної частини проекту, зарплати за розробку, оплати доставки та реклами.

Види стратегій залежно від ціни та якості продукції продемонстровані в таблиці 5.25.

Таблиця 5.25 – Види стратегій

Якість	Ціна		
	Висока	Середня	Низька
Висока	Висока ціна та якість товару	Зацікавлення покупця завдяки високій якості товару і середній ціні	Завоювання ринку, збільшення частку ринку
Середня	Мінімізуємо витрати на етапі впровадження	Встановлення середніх цін на товар середньої якості	Завоювання ринку, збільшення частку ринку
Низька	Несе загрозу втратити в майбутньому покупців	Несе загрозу втратити в майбутньому покупців	Встановлення низької ціни на товари низької якості

Наразі є план продавати понад 1000 преміум версій додатка у місяць в таблиці 5.26.

Таблиця 5.26 – План-продажів

Найменування показників	Періоди (по місяцях)												Всього за рік
	1	2	3	4	5	6	7	8	9	10	11	12	
Преміум версія додатку													
* обсяг ліцензій	10	20	30	40	50	70	90	100	120	150	150	170	1000
* ціна	\$ 500	\$ 1000	\$ 1500	\$ 2000	\$ 2500	\$ 3500	\$ 4500	\$ 5000	\$ 6000	\$ 7500	\$ 7500	\$ 8500	\$50000
* виручка	\$ 400	\$ 800	\$ 1200	\$ 1600	\$ 2000	\$ 2800	\$ 3600	\$ 4000	\$ 4800	\$ 6000	\$ 6000	\$ 6800	\$ 40000

Наразі цільовою аудиторією є малий та середній бізнес, проте в перспективі є заохочення великого бізнесу в різних галузях.

Для початку потрібно \$8000 для створення та підтримку сайту, мобільних додатків, юридичне оформлення.

Буде організована таргетна реклама в Інтернеті.

6. Виробничий план

- Купівля ліцензії на користування бази даних Oracle.
- Побудова інтерфейсу додатків та веб-сервісу на Angular JS.
- Масове.
- Додаткові функції для покращення роботи з користувачем за допомогою Штучного інтелекту.

7. Організаційний план

Профіль посади є дуже важливими при наймі працівників, зображені в таблиці 5.27.

Таблиця 5.27 – Профіль посади

Критерій	Зміст
Основна освіта	Вища освіта
Додаткова освіта, спеціалізація	Комп'ютерна спеціалізація
Необхідний досвід роботи	Щонайменше 2 рік
Завдання	Підтримка веб сайту та додатків
Знання	Oracle, Python, PHP, C#, Java Script, Android dev, Swift dev
Навички, вміння, ділові якості	Робота у команді
Особистісні якості	Вміння працювати з різними ПО
Мотивація (що можемо запропонувати)	Робота з дому

Стимулами будуть:

- матеріальні: заробітна плата; постійний оклад, премії за досягнуті показники ефективності;
- нематеріальні: можливість роботи, похвала, позиція в колективі та компанії, авторитет.

8. Фінансово-економічний план

На разі залучення інвесторів та власні кошти є орієнтованими джерелами фінансування нашого проекту таблиця 5.28.

Таблиця 5.28 – Планування загальних інвестицій по проекту

Група та вид інвестицій	Сума
Першопочаткові інвестиції, в т.ч.:	\$2000
Розробка веб-сервісу	\$10000
Розробка додатку під Android	\$4000
Розробка додатку під IOS	\$5000
Юридичні витрати	\$1000
Витрати на підтримку сайту та програм	\$5000
Витрати на рекламу	\$2000
ВСЬОГО ІНВЕСТИЦІЙ ПО ПРОЕКТУ:	\$29000

Заходи щодо упередження та реагування на ризики є в таблиці 5.29.

Таблиця 5.29 – Реагування на ризики

№	Ризик	Індикатори настання	Заходи щодо упередження	Заходи щодо реагування	Відповідальний
1	Некоректність роботи	Несприйнятність результатів	Відповідні оцінки моделювання	Вибір інших параметрів	Тертичний Роман
2	Зависання сайту	Припинення роботисервісу	Тестування	Вихід нових версій	Гуць Євгеній

Розрахунок ефективності проекту, що аналізується є важливою складовою в інвестиціюванні та наведено в таблиці 5.30.

Таблиця 5.30 – Розрахунок ефективності проекту.

Показник	2020	2021	2022	2023
1. Сума інвестицій	\$29000	-	-	-
2. Виручка від реалізації	-	\$18500	\$16000	\$25400
3. Витрати на експлуатацію проекту	-	\$3200	\$2400	\$2900
4. Амортизаційні відрахування	-	\$320	\$320	\$520
5. Ставка дисконту	-	18%	18%	18%
6. Грошові потоки	-	\$16000	\$14000	\$24000
7. Дисконтовані грошові потоки	-	\$13500	\$10000	\$14000
8. Дисконтовані грошовий потік з початку експлуатації проекту	-	\$13500	\$24000	\$38000
9. Дисконтовані грошові потоки за ставкою дисконту 40%	-	\$12000	\$7000	\$8500
10. Дисконтовані вигоди, тис. грн.	-	\$11500	\$12000	\$16000
11. Дисконтовані витрати, тис. грн.	-	\$3000	\$2000	\$19000

Розрахунок ефективності проекту

1. Сума інвестицій у проект становить \$29000.
2. Дисконтовані грошові потоки в результаті реалізації проекту становитимуть за 2020-2022pp. \$38000.
3. Чиста теперішня вартість проекту $\$38000 - \$29000 = \$9000$. Оскільки, $NPV > 0$, інвестиційний проект є вигідним для підприємства інвестора. За 3 роки функціонування проекту грошовий потік не лише задовольняє очікування інвестора у відношенні щодо одержання доходу, а й перевищують очікувані доходи на \$9000.
4. Термін окупності інвестицій. Застосуємо алгоритм розрахунку інвестицій:

$$TO = (t_0 - 1) + \frac{CI - \sum PV_{(t_0-1)}}{PV_{t_0}}$$

де t_0 - номер першого року, у якому досягається умова;

CI – сума інвестицій у проект;

PV – дисконтовані грошові потоки.

Отже, для даного прикладу, $TO = (3-1) + (29000 - 24000) / 24000 = 2,208$ роки

5. Внутрішня норма рентабельності. При ставці дисконту 40%, NPV дорівнює \$26000. Отже, $IRR = 0,18 + (38000 * (0,4 - 0,18) / (38000 - 26000)) = 0,89$, або при ставці 89% сумарні дисконтовані вигоди дорівнюють сумарним дисконтованим витратам. Тобто IRR є ставкою дисконту, при якій NPV проекту дорівнює нулю.

6. Коефіцієнт вигід/витрат. $\$40000 / \$7000 = 5,71$. За \$1 теперішньої вартості вкладених коштів у проект підприємство отримає \$5,71 теперішньої вартості доходу.

7. Індекс прибутковості $\$38000 / \$29000 = 1,31$ Отже, $PI > 0$ і проект є ефективним.

5.9 Інвестиційна та фінансове забезпечення стартап-проекту

Інвестиційна пропозиція

1. Назва проекту та відомості про організацію

Назва стартап-проекту: **FutureViews**

Галузева приналежність стартап-проекту: IT

Форма участі інвестора в стартап-проекті: надання інвестицій за договором простого товариства (договору про спільну діяльність)

Місце реалізації стартап-проекту: Gorizont1, м. Київ

Передбачувана дата початку реалізації стартап-проекту: 01.07.2020

Повне найменування організації / підприємства (партнера по проекту / ініціатора проекту): ТОВ “Ф’ючер В’юз”

Форма власності: приватна

Дата реєстрації: 01.04.2020

Поточні фінансові результати: -29000\$

Опис проекту (ідея/передумови, стан та етапи проекту, очікувані результати):

Ідея, передумови

FutureViews – веб додаток для завантаження, аналізу та прогнозування нестационарних, нелінійних, динамічних процесів у різних сферах застосування. Використовуючи статистичні дані, які нам необхідно проаналізувати користувач їх підвантажує та застосовує різноманітні математичні операції для аналізу та прогнозування.

Етапи:

1. Створення сайту, запуск маркетингової кампанії.
2. Створення мобільних додатків.
3. Розвиток сайту та додатків, розробка нового функціоналу.
4. Вихід на міжнародний ринок.

Результат - міжнародний проект масового форкасту, статистична платформа по набору даних

2. Характеристика ринку планованої до випуску продукції

2.1. Характеристика запланованої до випуску продукції:

- найменування та опис продукції: веб-сервіс futureviews.com, мобільний додаток FutureViews;
- основні споживачі: малий та середній бізнес у різних сферах діяльності та без кордонних меж;
- зацікавлені державні та науково-дослідницькі інститути у роботі з талановитою молоддю.

2.2. Опис поточного стану внутрішнього ринку планованої до випуску продукції:

- молодь України, від 15 до 35 років - 11 млн;
- кількість зацікавлених видів бізнесів більше 100 типів.

2.3. Основні показники внутрішнього ринку планованої до випуску продукції:

- річний обсяг промислового виробництва продукції, в поточних цінах: у 2021 році;
- рентабельність продажів продукції (в%): 45.7 (у 2021 році);
- ступінь концентрації (приблизна сумарна частка ринку п'яти найбільших гравців, в%): 100.

3. Перспективність і конкурентні переваги проекту

3.1. Опис наявної інфраструктури:

- Організації доставки Нова пошта, Укрпошта.

3.2. Стратегічні переваги:

- ненасичений ринок пропозиції;
- легкий вихід на міжнародну арену (організація доставки за межами України, підтримка серверів та трафіку сервісу).

4. Потреба у фінансуванні

Загальні інвестиційні витрати за стартап-проектом, всього: 29000\$, в тому числі капітальні витрати: 2200\$.

Потреба у фінансуванні проекту за рахунок інвестора: є, адже на розкрутку сервісу знадобиться багато часу, і потрібні будуть кошти на підтримку, доки сайт не почне приносити прибуток

5. Попередні фінансові показники проекту

Виручка, без ПДВ (після виходу на проектну потужність): \$24000

Простий /динамічний термін окупності: 2.2 роки

Внутрішня норма прибутковості (IRR): 89 %

Чиста поточна вартість (NPV): \$29000.

Ставка дисконтування: 18 %

Висновки до розділу 5

Результати даного розділу демонструють, ідею створення стартапу на дану тему є досить таки актуальна і цікава для сучасного суспільства і новітніх процесів в різних сферах її застосування.

Рентабельність усіх процесів під час створення та реалізації даного проекту повністю окупиться та зможе приносити прибутки, адже це актуальна проблема для ведення сучасної управлінської діяльності будь-якого виду керівництва. Функціонально-вартісний аналіз демонструє, що застосування вказаних інструментів має повністю покрити витрати на забезпечення якості програмного засобу та його реалізації.

Аналіз програмних середовищ та обчислювальні потужності кожного з ним тільки підтверджують те, що найкраще з такими задачами буде справлятися саме мова програмування Python.

ВИСНОВКИ ПО РОБОТІ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Представлена дисертація є результатом дослідження аналізу, моделювання та прогнозування часових рядів, для яких характерні нелінійність та нестационарність, а також розробленого програмного продукту за допомогою якого можна отримати емпіричні результати на основі реальних статистичних даних. У даній роботі було продемонстровані моделі, що побудовані на економічних даних – ріст ВВП США та екологічних даних – забруднення повітря.

У роботі розглянуто та досліджено ситуації так важливих чинників як стан економіки країн світу, вплив росту економіки на життя народу окремих держав та нестабільна й далеко не позитивна динаміка забруднення нашої екосистеми, що призводить до необоротних процесів та глобальної катастрофи людства.

У ході виконання дисертаційної роботи було створено набір різноманітних авторегресійних моделей та побудовано рекурентну нейронну мережу. Виконано аналіз випробувальних даних на наявність нелінійності, нестационарності, коінтегрованості, сезонності за допомогою відомих тестів. Зроблено графічну візуалізацію побудованих моделей на тестовій та валідаційній вибірках та в порівняльному режимі. Перевірено адекватність побудованих моделей на основі критеріїв адекватності та зроблено висновки щодо якості оцінювання прогнозу. Отримані результати та безпосередньо прогнозування на декілька кроків вперед імплементовано у графіки, що демонструють майбутню тенденцію характеру поведінки розглянутих часових рядів.

Що стосується результатів прогнозування росту економіки США, то можна було помітити, що вона має позитивну динаміку і продовжує рости,

хоча й має свої падіння. Проте кожна країна світу має прагнути досягти того ж рівня економіки на що повинно звернути увагу керівництво інших держав. Про екологічну ситуацію у світі такого позитиву сказати не можна, адже з кожним днем ситуація по забрудненню повітря стає все більше негативна. Звісно це можна апелювати тим, що вона має певну сезонність і вплив інших факторів, проте від цього наша планета не стає екологічно здоровішою. Тому потрібно усіма способами досягти пограження цих показників, адже за допомогою їхнього аналізу та прогнозування цього можна досягти, вивчаючи їхній характер.

Для покращення майбутніх досліджень економічних та екологічних процесів необхідно спробувати застосувати інші моделі регресійного аналізу та їх модифіковані типи, а також спробувати ансамбль цих методів. Для отримання більш цілісної картини можна застосувати пояснювальну змінну, що буде характеризувати вплив індексу споживчих цін на економічний ріст держави або ж рівень забруднення водоймищ на екологічний стан. Також доцільно буде розглянути інші методи моделювання та прогнозування такі як метод подібних траєкторій, баєсівських мереж та інші методи машинного навчання. Для покращення вибору моделі необхідно збільшити кількість рівнів перевірки адекватності моделювання та якості прогнозу, врахувати максимальну та мінімальну абсолютну похибку та метод максимальної правдоподібності.

ПЕРЕЛІК ПОСИЛАНЬ

1. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018. URL: <http://ghdx.healthdata.org/gbd-results-tool> (Last accessed: 14.11.2020)
2. Касперович С. А. Прогнозирование и планирование экономики : курс лекций для студентов специальностей 1-25 01 07 «Экономика и управление предприятием», 1-25 01 08 «Бухгалтерский учет, анализ и аудит», 1-26 02 02 «Менеджмент», 1-26 02 03 «Маркетинг». Минск. БГТУ. Минск: 2007. 172 с.
3. Антохонова И.В. Методы прогнозирования социально-экономических процессов: Учебное пособие. Улан-Удэ: Изд-во ВСГТУ. Улан-Удэ: 2004. 212 с.
4. Львович, М. И. Вода и жизнь. Водные ресурсы, их преобразование и охрана. Москва : Мысль, 1986. 254 с.
5. Athanasios Loukas, Exploring the Non-Stationary Effects of Forests and Developed Land within Watersheds on Biological Indicators of Streams Using Geographically-Weighted Regression; Seoul: Department of Environmental Planning, Konkuk University, 2016. 310 p.
6. AboutEViews: Part 2: Powerful Analytical Tools, Presentation Quality Output. URL: <https://www.eviews.com/EViews10/ev10analytics.html> (Last accessed: 15.11.2020)
7. Andrej Karpathy, The Unreasonable Effectiveness of Recurrent Neural Networks, Blog by Andrej Karpathy – 2015 URL:

<http://karpathy.github.io/2015/05/21/rnn-effectiveness/> (Last accessed: 15.11.2020)

8. Бідюк П.І., Коршевніук Л.О., Проектування комп'ютерних інформаційних систем підтримки прийняття рішень: Навчальний посібник. Київ: ННК «Інститут прикладного системного аналізу» Національний технічний університет України «Київський політехнічний інститут», 2010. 340 с.
9. Бідюк П.І., Половцев О.В. Аналіз та моделювання економічних процесів перехідного періоду. Київ: ПЛАБ-75, 1999. 230 с.
10. Бидюк П.И., Баклан И.В. Системный подход к построению регрессионной модели по временным рядам. *Системні дослідження та інформаційні технології*. 2002. № 3. С. 114-131.
11. Бідюк П.І. Часові ряди: моделювання та прогнозування. Київ: ЕКМО, 2004. 144 с.
12. Пашин В.П. Функционально-стоимостный анализ конструкторско-технологических решений. Київ: РДЭНТП «Знание» УССР, 1989. 222с.
13. Пашін В.П. Оцінка конкурентоспроможності електронних пристроїв на стадії проектування. Київ: Економічний вісник НТУУ „КПІ”, 2006. 255с.
14. Пашин В.П. Управление качеством изделий на основе функционально-стоимостного анализа. Київ: «Технология и организация производства», 1989. 177с.
15. Пашін В. П. Методичні вказівки до виконання економіко-організаційного розділу дипломних проектів (робіт) освітньо-

кваліфікаційних рівнів «бакалавр» і «спеціаліст» для студентів інституту прикладного системного аналізу. URL: <https://ela.kpi.ua/handle/123456789/1819> (Останній доступ: 18.11.2020р.)

16. Тертичний, Р. В. Моделі, методи та короткострокове прогнозування демографічних процесів в Україні : дипломна робота ... бакалавра : 6.050101 Комп'ютерні науки / НТУУ «КПІ ім. Ігоря Сікорського». Київ, 2019. 108 с.

ДОДАТОК А

ЛІСТИНГ ПРОГРАМИ

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
from statsmodels.tsa.seasonal import
seasonal_decompose
from statsmodels.tsa.stattools import adfuller, acf,
pacf, arma_order_select_ic
from statsmodels.graphics.tsaplots import plot_acf,
plot_pacf
from statsmodels.tsa.arima_model import ARMA
from statsmodels.tsa.arima_model import ARIMA
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
from statsmodels.stats.stattools import durbin_watson
from statsmodels.stats.diagnostic import
het_breuschpagan
from scipy.stats import shapiro
import warnings
warnings.filterwarnings("ignore")
df = pd.read_csv('GPD(USA).csv',
index_col='DATE', parse_dates=True)
df.index.freq = None
df.head(5)
df1 = pd.read_csv('GPD(USA).csv',
index_col='DATE', usecols = ['DATE', 'TimeSeries'],
parse_dates=True)
df.index.freq = None
df1.plot()
df["SMA1"] =
df["TimeSeries"].rolling(window=20).mean()
df["SMA2"] =
df["TimeSeries"].rolling(window=60).mean()
plt.figure(figsize=(20,10))
plt.plot(df["SMA1"], 'g--', label="SMA_20")
plt.plot(df["SMA2"], 'r--', label="SMA_50")
plt.plot(df["TimeSeries"], label="TimeSeries")
plt.legend()
plt.show()
df['middle_band'] =
df["TimeSeries"].rolling(window=30).mean()
df['upper_band'] =
df["TimeSeries"].rolling(window=30).mean() +
df["TimeSeries"].rolling(window=20).std()*2
df['lower_band'] =
df["TimeSeries"].rolling(window=30).mean() -
df["TimeSeries"].rolling(window=20).std()*2
plt.figure(figsize=(20,10))
plt.plot(df["TimeSeries"], label="Open")
plt.plot(df['upper_band'], 'g--', label="upper")
plt.plot(df['middle_band'], 'r--', label="middle")
plt.plot(df['lower_band'], 'y--', label="lower")
model = Sequential()
model.add(LSTM(units=50, return_sequences=True,
input_shape=(x_train.shape[1],1)))
model.add(Dropout(rate = 0.2))
model.add(LSTM(units=50, return_sequences =
True))
model.add(Dropout(rate = 0.2))
model.add(LSTM(units=50, return_sequences =
True))
model.add(Dropout(rate = 0.2))
model.add(LSTM(units=50, return_sequences =
False))
model.add(Dropout(rate = 0.2))
model.add(Dense(1))
# Compiling the model with **Stochastic Gradient
Descent** algorithm
# Using Mean Squared Error as Loss Function
model.compile(loss='mean_absolute_error',
optimizer='adam')
model.fit(x_train, y_train, epochs=100,
batch_size=64, verbose=1)
#predicting test data values, using past 90 from the
train data
inputs = all_data[len(all_data) - len(valid)-1:].values
inputs = inputs.reshape(-1,1)
inputs = scaler.transform(inputs)
inputs.shape
X_test = []
for i in range(1,inputs.shape[0]):
    X_test.append(inputs[i-1:i,0])
X_test = np.array(X_test)
X_test
X_test.shape
X_test = np.reshape(X_test,
(X_test.shape[0],X_test.shape[1],1))
preds = model.predict(X_test)
preds = scaler.inverse_transform(preds)
rms=np.sqrt(np.mean(np.power((valid-preds),2)))
#for plotting
train = all_data[:274]
valid = all_data[274:]
valid['Predictions'] = preds
plt.figure(figsize=(20,8))
plt.plot(train['pollution'])
plt.plot(valid['pollution'], color = 'blue', label = 'Real
pollution')
plt.plot(valid['Predictions'], color = 'red', label =
'Predicted pollution')
plt.title('Pollution air prediction')
plt.legend()
plt.show()
plt.legend()
plt.show()
plt.figure(figsize=(20,10))
plt.title('volume')
plt.plot(df['Vol'])
plt.show()

```

```

# ### Перевірка на стаціонарність за допомогою
# тесту Діккі-Фуллера
def adfuller_view(df):
    adf, p_value, used_lags, nobs, critical_values,
    max_info_criterion = adfuller(df)
    print('ADF =', adf)
    print('p-value =', p_value)
    print('Used lags:', used_lags)
    print('Number of observations:', nobs)
    print('Critical values:', critical_values)
    print('Maximized information criterion:',
    max_info_criterion)
    adfuller_view(df['TimeSeries'])
# Отримали значення p-value > 0.05. Отже,
# початковий ряд не є стаціонарним.
result_mult = seasonal_decompose(df['TimeSeries'],
model='multiplicative', freq=365)
plt.rcParams.update({'figure.figsize': (8,10)})
result_mult.plot()
plt.show()
df['TimeSeries'].pct_change()[1:].plot()
# ##### Можемо помітити, що наш ряд має
# лінійний тренд
# ARMA
df_log = np.log(df['TimeSeries'])
df_diff = df_log.diff(periods=1).dropna()
df_diff.plot(figsize=(20,10))
adfuller_view(df_diff)
# Бачимо, що p-value = 0, тому можна
# стверджувати, що ряд стаціонарний.
# ### Plot ACF and PACF
# p - порядок авторегресії, q - порядок ковзного
# середнього.
# Для визначення порядку авторегресії та
# ковзного середнього побудуємо графіки ACF та
# PACF.
fig = plt.figure(figsize=(12,8))
ax1 = fig.add_subplot(211)
fig =
sm.graphics.tsa.plot_acf(df1['TimeSeries'],lags=40,tit
le='ACF Non Stationary Data', ax=ax1)
ax2 = fig.add_subplot(212)
fig =
sm.graphics.tsa.plot_pacf(df1['TimeSeries'],lags=40,t
itle='PACF Non Stationary Data', ax=ax2)
plt.show()
plt.savefig('ACF_NonStationary_GPD(USA).png')
fig = plt.figure(figsize=(12,8))
ax1 = fig.add_subplot(211)
fig =
sm.graphics.tsa.plot_acf(df_diff,lags=40,title='ACF
Stationary Data', ax=ax1)
ax2 = fig.add_subplot(212)
fig =
sm.graphics.tsa.plot_pacf(df_diff,lags=40,title='PAC
F Stationary Data', ax=ax2)
plt.show()
plt.savefig('ACF_Stationary_GPD(USA).png')
AR = sm.tsa.ARMA(df_diff, order=(13, 0))
AR_fit = AR.fit()
print('Lag: %s' % AR_fit.k_ar)
print('Coefficients: %s' % AR_fit.params)

```

```

AR_fit.summary()
print ('durbin_watson score: %.4f'
%durbin_watson(AR_fit.resid))
# ### ARMA
ARMA_model = sm.tsa.ARMA(df_diff, order=(13,
6))
results_ARMA = ARMA_model.fit()
print ('RSS: %.4f' %
sum((results_ARMA.fittedvalues-
df_diff).dropna())**2))
plt.figure(figsize=(20,10))
plt.plot(df_diff)
plt.plot(results_ARMA.fittedvalues, color='red')
plt.show()
results_ARMA.summary()
predictions_ARMA_diff =
pd.Series(results_ARMA.fittedvalues, copy=True)
predictions_ARMA =
df_diff.add(predictions_ARMA_diff,fill_value=0)
predictions_ARMA =
df['TimeSeries'].add(predictions_ARMA,fill_value=0
)
print('RMSE: %.4f' %
np.sqrt(sum((predictions_ARMA-
df['TimeSeries'])**2)/len(df['TimeSeries'])))
print('RSS: %.4f' % sum((predictions_ARMA-
df['TimeSeries'].values)**2))
print ('r2_score: %.4f' %r2_score(df['TimeSeries'],
predictions_ARMA))
print ('durbin_watson score: %.4f'
%durbin_watson(results_ARMA.resid))
plt.figure(figsize=(20,10))
plt.plot(df['TimeSeries'], label='real')
plt.plot(predictions_ARMA, 'r+', label='predicted')
plt.legend(loc='best')
plt.show()
ARIMA_2_2_1 = sm.tsa.ARIMA(df_log, order=(2,
2, 1))
results_ARIMA = ARIMA_2_2_1.fit()
plt.figure(figsize=(20,10))
plt.plot(df_log.diff(periods=1).dropna().diff(periods=
1).dropna())
plt.plot(results_ARIMA.fittedvalues, color='red')
plt.show()
results_ARIMA.summary()
predictions_ARIMA_diff =
pd.Series(results_ARIMA.fittedvalues, copy=True)
predictions_ARIMA =
df_diff.add(predictions_ARIMA_diff,fill_value=0)
predictions_ARIMA =
df['Open'].add(predictions_ARIMA,fill_value=0)
print('RMSE: %.4f' %
np.sqrt(sum((predictions_ARIMA-
df['Open'])**2)/len(df['Open'])))
print('RSS: %.4f' % sum((predictions_ARIMA-
df['Open'])**2))
print ('r2_score: %.4f' %r2_score(df['Open'],
predictions_ARIMA))
print ('durbin_watson score: %.4f'
%durbin_watson(results_ARIMA.resid))
plt.figure(figsize=(20,10))
plt.plot(df['Open'], label='real')

```

```

plt.plot(predictions_ARIMA, 'r+', label='predicted')
plt.legend(loc='best')
plt.show()
from arch import arch_model
from pylab import rcParams
rcParams['figure.figsize'] = 20, 5
am = arch_model(results_ARIMA.resid,
mean='HARX', lags=100, vol='ARCH', p=2)
am_fit = am.fit(update_freq=0)
am_fit.summary()
fig1 =
sm.graphics.tsa.plot_acf(am_fit.resid.dropna().values,
lags=20)
fig2 =
sm.graphics.tsa.plot_pacf(am_fit.resid.dropna().value
s, lags=20)
plt.show()
# Оскільки значення лишків на графіках ACF та
PACF можна вважати близькими нулю та
значення коефіцієнта Акайке прямує до мінус
незкінечності, то можна вважати, що порядок
моделі підібраний вдало
print ('ARCH model')
print ('AIC: ', am_fit.aic)
print ('BIC: ', am_fit.bic)
print ('R-squared: ', am_fit.rsquared)
print ('durbin_watson score: %.4f'
% durbin_watson(am_fit.resid.dropna()))
# ARMA predict
ARMA_predictions = []
for i in range(len(valid)):
    model = sm.tsa.ARMA(train, order=(2, 1))
    model_fit = model.fit()
    output = model_fit.forecast()
    yhat = output[0]
    ARMA_predictions.append(yhat)
    obs = valid[i]
    train.append(obs)
    print('predicted=%f, expected=%f' % (yhat, obs))
ARMA_predictions = pd.Series(ARMA_predictions)
test_score = np.sqrt(mean_squared_error(valid,
list(ARMA_predictions)))
print('Test RMSE: %.5f' % test_score)
print ('r2_score: %.4f' % abs(r2_score(valid,
ARMA_predictions)))
print('RSS: %.4f' % sum((ARMA_predictions-
valid)**2))
# plot results
plt.figure(figsize=(20,10))
plt.plot(valid, label='Prices')
plt.plot(ARMA_predictions, color='red',
label='Prediction')
plt.legend()
plt.show()
predictions_ARMA_diff =
pd.Series(ARMA_predictions, copy=True)
predictions_ARMA = df_diff[-
10:].add(predictions_ARMA_diff,fill_value=0)
predictions_ARMA = df['Open'][-
10:].add(predictions_ARMA,fill_value=0)
plt.figure(figsize=(20,10))
plt.plot(df['Open'][-10:], label='real')
plt.plot(df['Open'][-10:], label='real')

```

```

plt.plot(predictions_ARMA[:10], 'r+', color='red',
label='Prediction')
plt.legend(loc='best')
plt.show()
# ARIMA predict
train_ARIMA = df_diff[:int(len(df_diff)-10)].tolist()
valid_ARIMA = df_diff[int(len(df_diff)-10):].tolist()
print('Train size: %d' % len(train))
print('Valid size: %d' % len(valid))
ARIMA_predictions = []
for i in range(len(valid_ARIMA)):
    model = sm.tsa.ARIMA(train_ARIMA, order=(2,
2, 1))
    model_fit = model.fit()
    output = model_fit.forecast()
    yhat = output[0]
    ARIMA_predictions.append(yhat)
    obs = valid_ARIMA[i]
    train_ARIMA.append(obs)
    print('predicted=%f, expected=%f' % (yhat, obs))
ARIMA_predictions =
pd.Series(ARIMA_predictions)
test_score =
np.sqrt(mean_squared_error(valid_ARIMA,
ARIMA_predictions))
print('Test RMSE: %.5f' % test_score)
print ('r2_score: %.4f' % r2_score(valid_ARIMA,
ARIMA_predictions))
print('RSS: %.4f' % sum((ARIMA_predictions-
valid_ARIMA)**2))
# plot results
plt.figure(figsize=(25,8))
plt.plot(valid_ARIMA, label='Prices')
plt.plot(ARIMA_predictions, color='red',
label='Prediction')
plt.legend()
plt.show()
predictions_ARIMA_diff =
pd.Series(ARIMA_predictions, copy=True)
predictions_ARIMA = df_diff[-
10:].add(predictions_ARIMA_diff,fill_value=0)
predictions_ARIMA = df['Open'][-
10:].add(predictions_ARIMA,fill_value=0)
plt.figure(figsize=(20,10))
plt.plot(df['Open'][-10:], label='real')
plt.plot(predictions_ARIMA[:10], 'r+', color='red',
label='Prediction')
plt.legend(loc='best')
plt.show()
import pandas as pd
import numpy as np
import datetime as dt
from datetime import datetime
import matplotlib.pyplot as plt
import numpy as np
from sklearn.preprocessing import MinMaxScaler
### Create the Stacked LSTM model
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import LSTM
from keras.layers import Dropout
#read the file

```

```

df = pd.read_csv("air_pollution_2014.csv")
#print the head
df.head()
# Setting Date as Index. We will use **Close** Price
for prediction
#setting index as date
df['date'] = pd.to_datetime(df.date)
df.index = df['date']
#plot
plt.figure(figsize=(20,8))
plt.plot(df['pollution'], label='pollution')
df.describe()
# Selecting Close Price as feature and removing other
columns from the data
features = ["date", "pollution"]
all_data = df[features]
#setting index
all_data.index = all_data.date
all_data.drop('date', axis=1, inplace=True)
all_data.head()
# In[11]:
all_data.shape
#creating training and validation sets
dataset = all_data.values
train = dataset[:274,:]
valid = dataset[274:,:]
valid.shape
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(dataset)
# This data structure is needed to cover 90-days close
price stamps, based on which RNN will predict the
91st day's close price.
x_train, y_train = [], []
for i in range(1,len(train)):
    x_train.append(scaled_data[i-1:i,0])
    y_train.append(scaled_data[i,0])
x_train, y_train = np.array(x_train), np.array(y_train)
x_train = np.reshape(x_train,
(x_train.shape[0],x_train.shape[1],1))
# Initializing the LSTM model and 2nd, 3rd and 4th
LSTM layer each with a Dropout Layer. The layers
contain 50 neurons and with a Dropout rate of 20%,
twenty percent of 50 neurons will be ignored
randomly during each iteration
# Finally, an output layer is added with 1 as an output
dimension (as we are predicting the close price)
# create and fit the LSTM network

```

